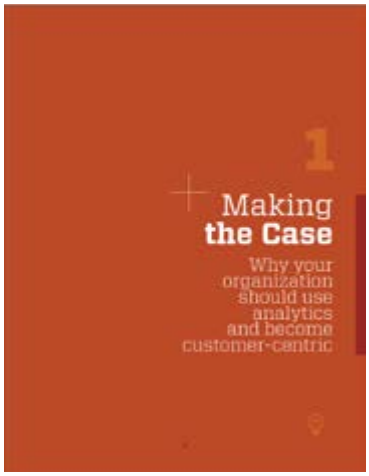
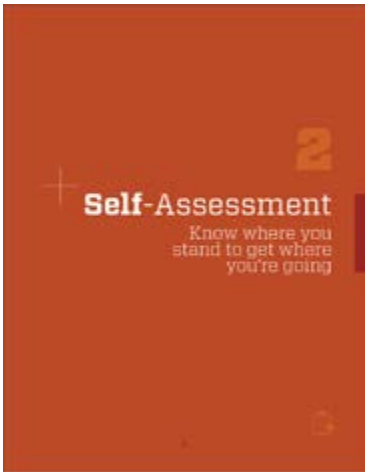
Customer Analytics Toolkit

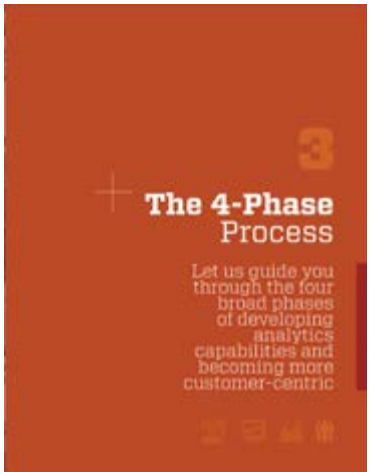
DRIVING CUSTOMER CENTRICITY
THROUGH ANALYTICS

CGAP

How to Use **This Toolkit**







Making the Case.

Customer centricity is about providing solutions based on a deep understanding of customer needs, preferences, and behaviors.

Learn why your organization should use analytics and become more customer-centric.

Self-Assessment.

Effective analytics that support customer centricity require organizational leadership and attitude, data capabilities, and analytical capabilities.

Rate your organization on these three capabilities to know where you stand.

The 4-Phase Process.

Four broad phases guide the customer analytics process – from initial visioning to creating final outputs for dissemination and engagement.

Follow these four phases to develop your analytics capabilities and become more customer-centric.

Table of Contents

Making the Case ..... p. 4

What Is Customer Centricity and How Can Analytics Drive It? ..... p. 7

Self-Assessment ..... p. 8

The 3 Mutually Supportive Capabilities ..... p. 10

Organizational Leadership and Attitude ..... p. 11

Data Capabilities ..... p. 14

Analytical Capabilities ..... p. 17

How Effective Is Your Organization? ..... p. 19

The 4-Phase Process ..... p. 20

Phase 1: Visioning ..... p. 22

Phase 2: Assess Possibilities ..... p. 36

Phase 3: Refine Plan, Data Request, and Analysis ..... p. 62

Phase 4: Dissemination and Engagement ..... p. 92

Case Study ..... p. 100

Card Pioneer Microinsurance Data Analysis Project ..... p. 100

COVER PHOTO: Working on a Sewing Machine, India — Photo by Rana Pandey

A man works on a sewing machine to meet his daily target. He works alongside many others in a small, dimly lit room.

Home Sweet Home, Myanmar — Photo by Khant Zaw

Minn village is a two hour drive from Yangon City. Its people make their living by catching, drying, and selling fish. As I walked through the village during my visit, a fisherman's family invited me to lunch in their small bamboo hut. They were friendly and it was fun.

1

Making **the Case**

Why your organization should use analytics and become customer-centric

Fisherman and Big Nets, Indonesia — Photo by Amrullah Basri Paembonan

A fisherman uses big nets in a coastal area of Binongko Island. Fishermen on the island are classified as poor and fishing is just one part of their livelihood. On average, their income is no more than $20 a month.

What Is Customer Centricity and **How Can Analytics Drive It?**

Customer centricity is about providing solutions based on a deep understanding of customer needs, preferences, and behaviors.

Financial service providers generate vast amounts of customer data, such as demographics data collected at enrollment and behavioral data collected as customers transact and use products and services. This data is extremely helpful for a better understanding of who customers are, what their needs are, the problems they experience, and how they feel about an organization. Used effectively, data creates competitive advantage and helps organizations better serve their customers.

This is especially important in financial inclusion where a focus on lower costs may cause financial service providers to lose sight of customer experience. There may be a significant gap between the lives and experiences of lower income customers and the management teams, operational teams that design products, and front-line staff that serve them (and may not easily relate to or understand their needs, preferences, and behaviors).

For more on customer centricity, read the CGAP brief, Customer Centricity for Financial Inclusion

Camera Mechanic, India — Photo by Arabinda Dutta

This 70-year-old man was a master mechanic in the era of analog photography. He started his career in a small rented shop in central Kolkata. Now, his challenge as a poor, less-educated mechanic is to tap into fast-growing digital technology. He took small bank loans to purchase an older digital camera since he knows he has to survive in his profession and feed his family.

2

Self-Assessment

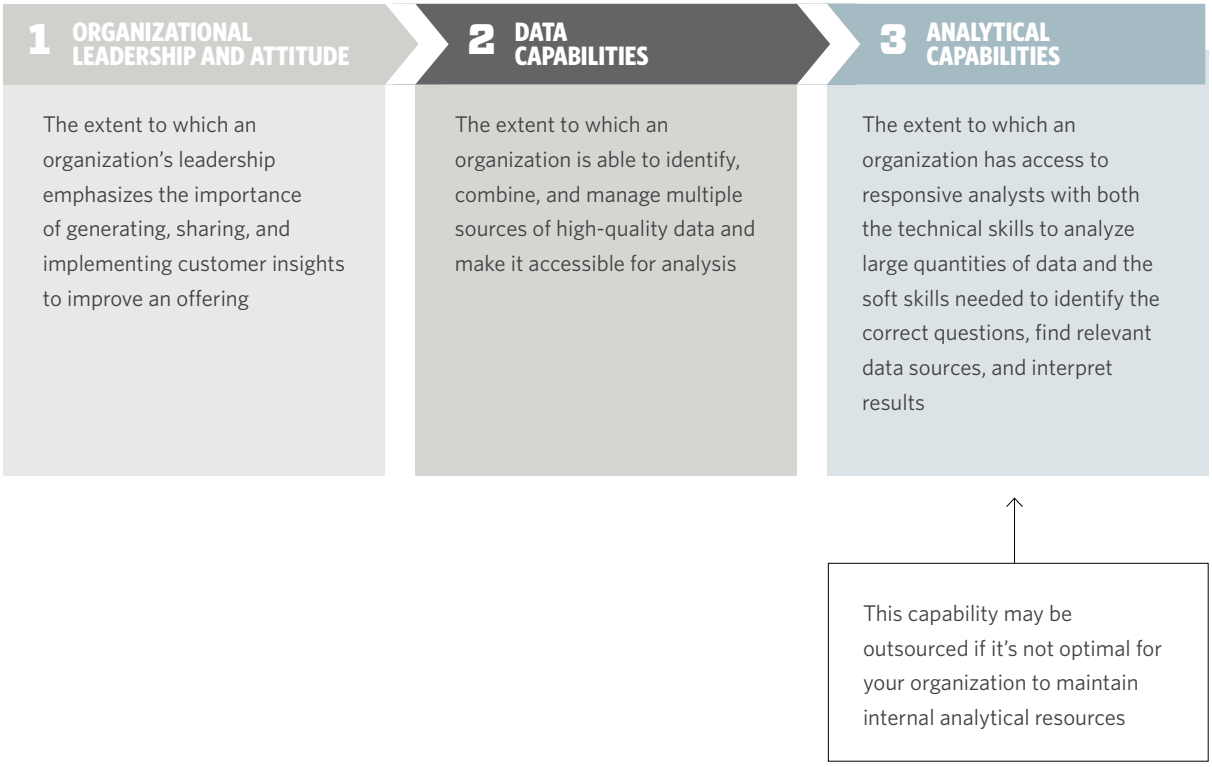
Know where you stand to get where you're going

The 3 **Mutually Supportive Capabilities**

This toolkit targets organizations that are starting their journey toward becoming more customer-centric through analytics. It's designed to help financial service providers optimize their use of INTERNAL DATA to better understand customer behavior and generate insights.

A key complexity is the diversity of financial service providers and where they currently stand on the journey toward customer centricity. Effective analytics require three mutually supportive capabilities:

Mutually Supportive Capabilities



1 | Organizational Leadership and Attitude

KEY CHALLENGES

Let's look at each capability. Here are some key pitfalls linked to organizational leadership and attitude.

Insights Culture

Insufficient buy-in at the executive level

Insights Sharing and Usage

Limited mechanisms to share learnings across the organization so customer-centric insights frame both operational and strategic decision-making

Targets and Performance Measures

Metrics and targets focus on product, short term profitability, and sales

Experimentation

Limited willingness to act on insights generated through analytics by directly testing and experimenting solutions with customers

Quiz

Explore these dimensions to further characterize your organization. Score yourself on each subcategory and rate your organization on page 19, *How Effective Is Your Organization?*

INSIGHTS CULTURE A level of empathy with customers runs through the entire organization. The extent to which leadership emphasizes the importance of generating customer insights to better serve the market	There's broad recognition that our organization is not currently customer-centric	Our organization is starting to talk about customer centricity, but we recognize that there's a lot of room for improvement	Our organization self-identifies as being customer-centric, although we're always looking for ways to improve
ANSWER	LOW SCORE: 1 POINT	MEDIUM SCORE: 2 POINTS	HIGH SCORE: 3 POINTS
	☐	☐	☐

INSIGHTS SHARING AND USAGE Extent to which customer data is shared throughout the organization and used by all functions and departments	Insights are not effectively shared across our entire organization. There's limited or zero usage of insights within the strategic and product development functions	We're improving the way insights are shared across the entire organization. Multiple divisions are encouraged to incorporate insights into strategy and product development	Insights are actively shared across our entire organization and used by multiple divisions, including strategic and product development functions
ANSWER	LOW SCORE: 1 POINT	MEDIUM SCORE: 2 POINTS	HIGH SCORE: 3 POINTS
	☐	☐	☐

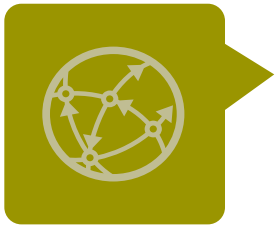
TARGETS AND PERFORMANCE MEASURES Extent to which targets and reporting focus on long-term customer lifetime value (customer-driven metrics) rather than short-term sales and profitability	We do not make use of customer metrics. Our targets are driven by sales – and business performance is measured in terms of sales	Our reporting includes some customer metrics but the majority still focuses on profitability and sales metrics	Our ongoing reporting incorporates key customer-related and engagement metrics. Our targets and performance measures focus on long-term customer lifetime value rather than short-term sales
ANSWER	LOW SCORE: 1 POINT	MEDIUM SCORE: 2 POINTS	HIGH SCORE: 3 POINTS
	☐	☐	☐

EXPERIMENTATION Extent to which the organization is willing to experiment with the customer value proposition, service design and optimization, to try and fail, and to incorporate learnings	We currently do not use experimentation in product design and development	Our use of experimentation is very limited	We use experimentation in product design, development, and optimization
ANSWER	LOW SCORE: 1 POINT	MEDIUM SCORE: 2 POINTS	HIGH SCORE: 3 POINTS
	☐	☐	☐

Total Score ☐

2 Data Capabilities

KEY CHALLENGES
Here are some key pitfalls related to data capabilities.



Data Sources/ Collection

Limited focus on collecting customer data because data has not been considered a business asset

In some cases, too much data is collected and the organization is unable to manage the quantity generated or prioritize what's useful

Data Quality

Data collected is often not standardized or verified because it's seldom referenced

Single Customer View

Data sources are separate and difficult to connect, so the organization has a fragmented view of their customers

Quiz

DATA SOURCES/ COLLECTION Extent to which the organization actively gathers useful customer data from multiple touchpoints, including: Internal/external data Primary/secondary data Qualitative/quantitative data	We largely focus on easy-to-access external secondary data	We make extensive use of external secondary data and primary qualitative data, but we currently make limited use of internal business data	We actively gather customer data (both qualitative and quantitative) from multiple touchpoints across our organization. We often use external data to fill gaps or overlay onto internal data
ANSWER	LOW SCORE: 1 POINT	MEDIUM SCORE: 2 POINTS	HIGH SCORE: 3 POINTS
	☐	☐	☐

DATA QUALITY Extent to which data is in an easily accessible environment – and of high quality	We currently have no data warehouse and fully rely on source systems to access internal data. The data is not clean and often only partially populated	We're starting to more efficiently collect and maintain customer data, with a focus on improving accessibility and quality	Our data is largely complete, of high quality, and easily accessible from a data warehouse
ANSWER	LOW SCORE: 1 POINT	MEDIUM SCORE: 2 POINTS	HIGH SCORE: 3 POINTS
	☐	☐	☐

Street Vendor, India
— Partha Sarathi Dalal, CGAP Photo Contest
A Varanasi street vendor sells tomatoes door-to-door in order to support his family.

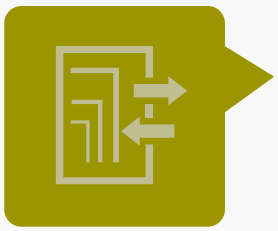
SINGLE CUSTOMER VIEW Extent to which the organization is able to see a single customer view	We currently do not have a single customer view	We are in the process of creating a single customer view. Some databases have been linked but a lot needs to be done	We are able to see all data across various sources for a single customer
ANSWER	LOW SCORE: 1 POINT	MEDIUM SCORE: 2 POINTS	HIGH SCORE: 3 POINTS
	☐	☐	☐

Total Score ☐

3 | Analytical Capabilities

KEY CHALLENGES

Here are some key pitfalls related to analytical capabilities.



Analysis Capabilities

Limited access to resources that have the technical capabilities to analyze customer data

Analysts are unfamiliar with a customer-centric way of thinking and may lack the creativity or the ability to determine key questions about customers (and how best to answer them)

Data Reporting Expertise

Ineffective reporting on findings means customer insights are engaged with ineffectively

Quiz

ANALYSIS CAPABILITIES Extent to which the organization has responsive analysts who generate rich insights in response to business needs	We currently have limited internal capabilities to access and analyze data. To create customer value propositions, our analysis principally focuses on secondary data that's widely available in the market	We feel comfortable with qualitative data, however, we have limited internal capabilities to efficiently analyze quantitative internal data	We have responsive analysts who generate rich insights in response to organizational needs. They're given time to engage with data and generate interesting insights
ANSWER	LOW SCORE: 1 POINT	MEDIUM SCORE: 2 POINTS	HIGH SCORE: 3 POINTS
	☐	☐	☐

DATA REPORTING EXPERTISE Extent to which analysts are able to generate and communicate useful, effective insights that encourage engagement throughout the entire organization	We do almost no internal reporting on customers	We do some internal reporting, but currently it's not effectively communicated throughout the organization and there's room for improvement	We internally generate useful and engaging reports on customers that are distributed throughout the organization
ANSWER	LOW SCORE: 1 POINT	MEDIUM SCORE: 2 POINTS	HIGH SCORE: 3 POINTS
	☐	☐	☐

Total Score ☐

How Effective **Is Your Organization?**

How does your organization score on each capability?

	AT THE BEGINNING	ON ROUTE	ARRIVED
ORGANIZATIONAL LEADERSHIP AND ATTITUDE	☐ 0-4 POINTS	☐ 5-7 POINTS	☐ 8 OR MORE POINTS
DATA CAPABILITIES	☐ 0-4 POINTS	☐ 5-7 POINTS	☐ 8 OR MORE POINTS
ANALYTICAL CAPABILITIES	☐ 0-4 POINTS	☐ 5-7 POINTS	☐ 8 OR MORE POINTS

Where do you **need help?**

ANSWER

Vegetable Vendor of Leh, Ladakh — Photo by Anindya Majumdar

This traditional vegetable market has existed on the pavement of Leh's main market for decades. The vendors are women who grow vegetables in nearby villages and carry them to sell at the market. The money they earn is solely their own. They're determined to guard their economic independence and have successfully resisted the administration's repeated attempts to close the market in the name of urbanization.

3

The 4-Phase Process

Let us guide you through the four broad phases of developing analytics capabilities and becoming more customer-centric

VISIONING
ASSESS POSSIBILITIES
REFINE PLAN/DATA REQUEST/ANALYSIS
DISSEMINATION/ENGAGEMENT

Phase 1
Visioning

Phase 1
Visioning

Part A
Who's going to lead the initiative?

Part B
What do we want to know about our customers?

Phase 2
Assess Possibilities

Part A
What's possible given the data?

Part B
Do we have a single customer view?

Phase 3
Refine Plan, Data Request, and Analysis

Part A
How do we obtain and prepare data for analysis?

Part B
How can we effectively conduct the analysis?

Phase 4
Dissemination and Engagement

Part A
What kinds of outputs work best?

Part B
Return to Phase 1: Feed back into the ongoing visioning process

PART A

Who's going to **lead the initiative?**

It's critical to identify a "customer insights champion" to drive the process of generating insights from analytics. This person acts as your team lead and manages the process.

THE TEAM LEAD/CUSTOMER INSIGHTS CHAMPION

- Champions the process internally, ensuring that the right resources are allocated to the project and that outputs and deliverables are used across the organization
- Ideally is **curious about customers, able to identify interesting questions and creatively identify the best available data to answer them, and can manage various teams** in different areas of the organization

MAJOR RESPONSIBILITIES

With the support of business leadership, responsibilities include:

- Identifying key questions related to customers and use cases for insights
- Identifying potential project resources, including people who work with data and necessary analytical resources (may be internal or external analytics people)
- Ensuring that data and analytical resources effectively communicate and the project is a priority for all involved
- Consistently providing leadership with feedback on progress and outcomes
- Ensuring that insights generated are shared with relevant people across the organization and actioned to improve how customers are served
- Monitoring initiatives and making adjustments as required

Life of a Flower Seller, India — Aniket Mazumder, CGAP Photo Contest

Life is not a bed of roses for flower sellers in India. This portrait stuck in my heart as I came to know that a cot under a bridge is where this man lives with his wife and children.

PART B

What do we want **to know about our customers?**

It's useful to think about what you actually want to know about your customers up front, how this fits with your business strategy, and how data can drive the process.

It's likely that conversations about customer centricity and how data can support it have happened throughout your organization. However, it's beneficial to set up an initial workshop with key stakeholders and resources to discuss business challenges and which questions about customers you'd like to answer.

BESIDE THE TEAM LEAD, WHO SHOULD BE INVOLVED?

During the visioning phase, it's beneficial to involve people from different areas of your organization. They may or may not be a part of the process, however, it's useful to involve at least one person from the following areas:

BUSINESS LEADERSHIP

- Drives the process and ensures that generating customer insights from data is seen as an organizational priority
- Ensures that insights are shared with relevant people across the organization and actioned to improve how customers are served

DATA TEAM

- Has the technical knowledge and skills required to manage multiple data sources and extract the necessary data for analysis. Has the best understanding of which data is available, its quality, and how easy it is to access
- May consist of one or more of the following: IT people, database administrators, data analysts, developers

ANALYTICS TEAM

- Helps shape which customer insights are generated
- While business leadership has a vision of what they want to know about customers, the analytics team assesses what's possible given available data, runs the actual analysis, interprets results, and communicates them back to the organization

Key Considerations **for Your Visioning Workshop**

A clear vision of the desired business impact will help shape the process going forward in terms sourcing data, running the actual analysis, and framing how insights are communicated back to your organization.

CONSIDER INSIGHTS AND METRICS – AS WELL AS POTENTIAL OUTPUTS

- Having an idea of what you want to know about your customers and what potential outputs may look like helps shape your analysis and ensures its usefulness
- The possibilities are endless. Analytics aid operational interventions, such as simple customer reporting that empowers front-line staff and managers to better serve customers. More strategic interventions identify potentially profitable customers and how to better serve them, i.e., finding emerging trends within your customer base.

REMAIN OPEN TO NEW AVENUES OF EXPLORATION

- While a clear objective is important, engagement with data almost certainly leads to new questions, unexpected insights, and new avenues of exploration
- Analytics is an iterative process and you may need to adjust the vision over time

INITIATE A DISCUSSION ABOUT DATA

- Start by reviewing which data is currently available and where gaps exist
- Include internal data as well as qualitative data (i.e., discussions with customers) that you may have access to. If qualitative data is available, review it first since it may highlight key focus areas for customer analytics

START SMALL

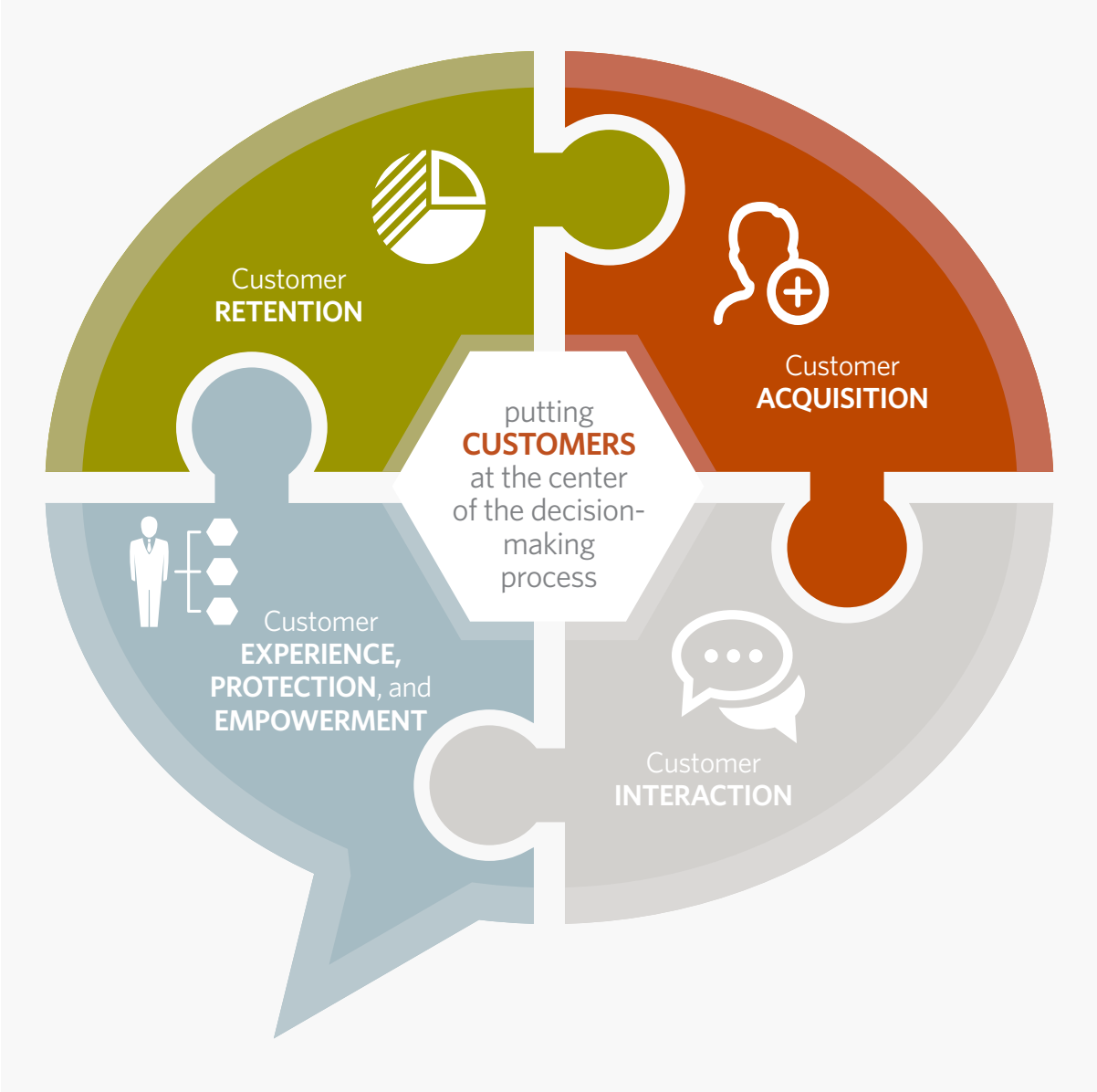
- Simple customer metrics and analytics, such as aggregating and summarizing customer data, can have a strong business impact and spark new, more complex questions. There's a much higher chance of success if you start with manageable tasks
- Basic analysis also helps you explore existing data. In some cases, management may believe that data is better, or more accessible, than it actually is

STAY FOCUSED ON CUSTOMERS

- It's easy to get sidetracked and shift your focus from customer-centric to product-centric analytics. Be sure to stay focused on customers

Understanding **Customers**

Given your business strategy and objectives, what do you want to understand about customers?



Here are potential questions your organization may have about customers. They can be adjusted based on the particular business challenges being addressed.

Customer **ACQUISITION**

Who are our new customers?

Through which channels did they come to purchase our products?

Did they engage with a sales campaign?

What's our cost to aquire new customers?

How can we improve the acquisition process?

Customer **INTERACTION**

Do different customer groups/segments behave differently?

What are preferred transaction and servicing channels for customers?

Do customers read or respond to various communications?

Which customers are more likely to go into arrears?

Which customers use multiple products?

How can we better serve different customer segments?

Customer **EXPERIENCE, PROTECTION, and EMPOWERMENT**

How satisfied are our customers with our current offerings?

Why are they satisfied? What are we doing well? What can we improve?

Are sales practices in the best interest of our customers?

Are we selling the correct products to the right customers?

Customer **RETENTION**

How actively do customers use our products?

What are inactivity or dormancy rates?

Which customers use a service once – then never use it again?

Which customer/ customer groups are terminating?

Why are they terminating?

How can we reduce churn?

The Making, India — Nivedita Dutta, CGAP Photo Contest

When glaze meets an optimal melting point it hardens. This man stacks dry clay pots that will be fired and turned into the final product.

Examples of **Customer Metrics**

The following metrics can be tracked over time and looked at by customer demographics (age, gender, area, etc.) and/or customer segment (if segmentation has been done).

CUSTOMER LIFETIME VALUE

- Predicted net profit attributed to the entire future relationship with a customer or customer group

MARKETING

- Customer reach by marketing/communications channel and campaigns
- Customers signed up for loyalty programs
- Customer engagement with loyalty programs (points accumulation, redemption, etc.)

PRODUCT

- Number of customers with each product/service
- Number of customers with multiple products (cross-holding)

COMPLAINTS AND DISPUTES HANDLING

- Number of customer complains
- Number of customers responded to
- Number of customers with complains resolved or escalated
- Number of customers satisfied with the outcome

DELINQUENCY MANAGEMENT

- Number of customers who missed a payment or are in arrears

CUSTOMER SATISFACTION

- Customer satisfaction score
- Repeat purchase intention
- Net Promoter Score

SALES/UPTAKE

- New and repeat customers
- Number of customers who engage with sales campaigns
- Cost of customer acquisition

CHURN

- Number of customers who've terminated

TRANSACTING/ACTIVITY

- Customer activity and dormancy rates
- Customers' preferred transaction channels
- Customers by size and type of transaction

SERVICING/SUPPORT

- Number of customer queries
- Average reply and resolution time
- Call wait time and abandonment rate
- Number of customers satisfied with servicing and support

The Responsibility, India — Sujan Sarkar, CGAP Photo Contest

A woman's responsibilities often include household chores as well as agricultural practices.

Which Metrics to Use – **and Why**

DEPENDS ON YOUR ORGANIZATION'S DATA AND ANALYTICAL CAPABILITIES

- For organizations just beginning to use analytics to become more customer-centric, simple metrics are the best place to start. Try, for example: number of new customers, number of customer queries, number of customers signed up for each product or service
- As your data and analytical capabilities grow more complex, metrics such as customer lifetime value can be tracked by specific customer segment

DEPENDS ON YOUR ORGANIZATION'S KEY FOCUS AREAS

- Credit organizations, for example, should carefully track delinquency, while organizations that struggle with churn should track the number of customers who've terminated over the past week/month – and carefully monitor complaints, customer satisfaction, and service and support

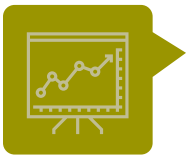
More on metrics in Phase 3, Part B: *How can we effectively conduct the analysis?*

Phase 2
Assess Possibilities

Phase 1
Visioning

Part A
Who's going to lead the initiative?

Part B
What do we want to know about our customers?



Phase 2
Assess Possibilities

Part A
What's possible given the data?

Part B
Do we have a single customer view?

Phase 3
Refine Plan, Data Request, and Analysis

Part A
How do we obtain and prepare data for analysis?

Part B
How can we effectively conduct the analysis?

Phase 4
Dissemination and Engagement

Part A
What kinds of outputs work best?

Part B
Return to Phase 1: Feed back into the ongoing visioning process

PART A

What's possible **given the data?**

The objective of Phase 2 is to gain an understanding of which customer data is available for analysis.

HERE ARE THREE KEY QUESTIONS TO ASK:

1. Which data do we currently collect?

2. Which data is available for analysis?

3. Where's the data stored?

Mother Fishmonger, Indonesia — Thaib Chaidar, CGAP Photo Contest

Portrait of a fish trader at a traditional market in Malino, South Sulawesi.

Question 1:
Which data **do we currently collect?**

Before getting into the technicalities of what data sits where, it's useful to map out the data collection process.

- Organizations often believe they know what data they're collecting – only to discover they're not collecting as much as they thought or are collecting a large amount that's not useful
- A good understanding of what's collected helps optimize the data collection process by indicating gaps and what may be collected but not utilized. The process also provides useful context for analysis

Example: An Insurance Company's Data Collection Process

In an analysis of customer data for a large insurance company, the analytics team found that enrollment data for certain customers was of higher quality and more detailed than for others. A closer look at the enrollment process indicated that different data was collected for different sign-up channels (branch vs. independent agents). The data mapping process examined the different enrollment touchpoints and which data was collected. This provided context for the data team and indicated that the organization should optimize and standardize data collection at all enrollment channels.

HOW TO DO THIS?

Map out all customer touchpoints and document which data is collected at each.
See *The Data Mapping Process* on the following page.

The Data **Mapping Process**

Customer Acquisition

KEY QUESTIONS:

- What are the touchpoints where customers can acquire our products and services?
- Which data is collected at each touchpoint?
- If partners are involved, can we see all the data they collect?
- What's the **quality** of the data collected? (Quality is impacted by the completeness and veracity of data customers provide and how accurately it's captured)

Servicing

KEY QUESTIONS:

- Which servicing channels do customers use, including servicing and support, transacting, claims, marketing and communication touchpoints?
- If partners are involved, can we see all the data they collect?
- What's the **quality** of the data collected?

How can we optimize data collection?

- Can we improve data quality? For example, if address data is collected but currently not usable because it's in free text format (i.e., customers input manually in a non-standardized format), consider implementing a process that asks customers to select from a drop-down list
- Can data collected be standardized across touchpoints?
- Can less useful data be dropped or replaced by higher quality data collected by a more useful question?

This could form part of a broader customer journey mapping process:

The CGAP Customer Journey Map

Wheel in the Industry, India — KM Asad, CGAP Photo Contest

A man works as part of the Kolkata garment industry at a small hosiery thread factory near his house.

The data mapping process **should account for all internal data sources**

Examples of internal data sources include:

PRODUCTION DATA

- Transactions data generated through various channel points: agents, branches, ATMs, POS, internet, etc.

ACCOUNTS DATA

- Snapshot of account status, balance, contractual status, etc.

CAMPAIGN DATA

- Data on promotions that have run
- ID on customers notified about a promotion

COMMUNICATIONS DATA

- Inbound and outbound sales communications, including calls, SMSs, emails, etc.
- Service communication: Agent collects ID numbers from customers and links to them

COMPLIANCE DATA

- Data collected for regulatory compliance such as anti-money laundering and Know Your Customer

LOYALTY DATA

- Coupons or vouchers issued
- Usually links to transaction data

APPLICATION DATA

- Data collected on enrollment: age, address, income, employment, gender, etc.

CUSTOMER SATISFACTION DATA

- Surveys conducted with customers
- Focus groups with customers

BRANCH DATA

- Often not maintained
- May include a geographic indicator

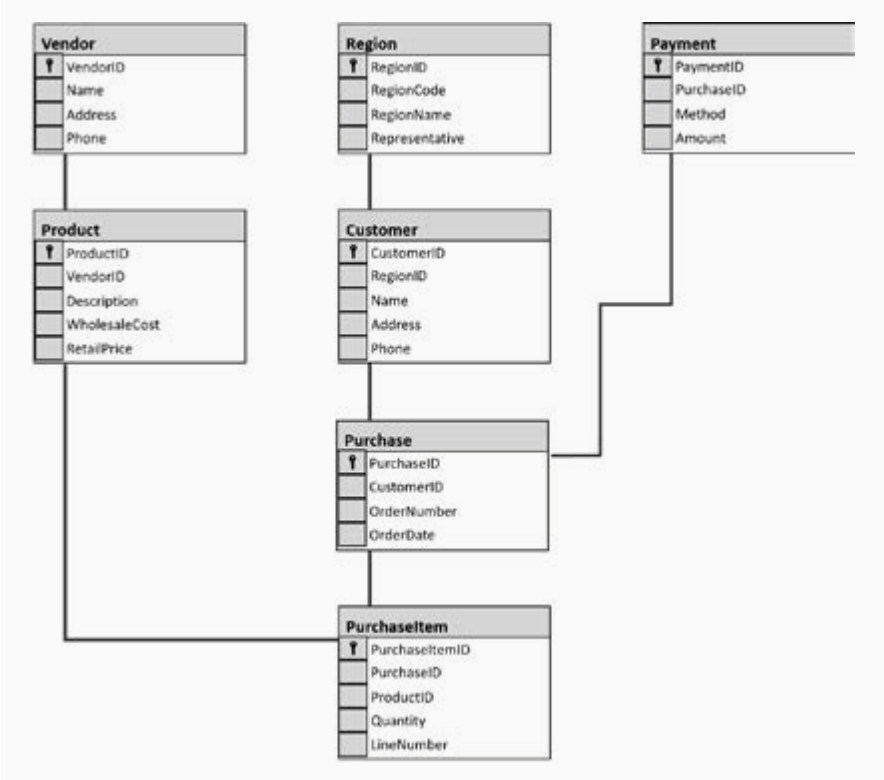
Question 2:
Which data is **available for analysis?**

Once you've mapped out the data collection process, it's time to assess which data is available for analysis. Database architecture and a data dictionary are useful for understanding database structure and available data in more detail.

DATABASE ARCHITECTURE

- A set of specifications, rules, and processes that dictate how data is stored and accessed by components of a system
- Includes data types, relationships, and naming conventions
- Specifies primary keys (a field in a table whose value uniquely identifies each record)
- Describes the organization of all database objects and how they work together

Database architecture maps out the various data fields available and how databases relate to each other



DATA DICTIONARY

- A centralized repository of information about data that includes meaning, origin, usage, format, and relationship to other data
- Importantly, includes a description of each field. For example, if there's a field called "claim data," is it the date a claim was made, processed, or paid out?
- A description of codes, if applicable. For example, a loan status field with a key that describes the various statuses
- Because data may be adjusted over time, it's important to have someone maintain the data dictionary and keep information current

A **data dictionary** is used to control and maintain databases.

Field Name	Data Type	Field Length	Constraint	Description
Client_id	Int	10	Primary	Client ID, Auto generated
Client_name	Varchar	20	Not null	Name of client
Password	Varchar2	30	Not null	Login Password for client
Contact_no	Int	15	Not null	Landline or mobile number
Email_id	Varchar2	30	Not null	Any email ID
Max_Users	Int	10	Not null	Maximum number of users
Current_users	Int	10	Not null	Currently present user

If your organization does not have a data dictionary, it's worth creating one and appointing someone to maintain and update it.

Question 3: Where's the **data stored?**

Another key question for Phase 2 is where does data sit: in a source system or in a data warehouse?

SOURCE SYSTEMS

- Also known as operational systems. Designed to support typical day-to-day operations via individual user transactions
- Collect customer data at the source
- Built to deal with large volumes of detailed data
- Generally process-driven, meaning focused on specific organizational processes or tasks like billing, registration, etc.
- Generally concerned with current data and usually updated regularly so data is frequently overwritten
- Can reveal a snapshot of ongoing organizational processes but not designed to support decision-making

DATA WAREHOUSE

- Acts as a centralized repository of an organization's data
- Primary goal is to enable users appropriate access to a homogenized, comprehensive view of the organization, supporting forecasting and decision-making processes at the enterprise level
- An additional goal is to achieve informational consistency by bringing data from disparate sources into a centralized repository that provides a homogenized view of the organization's data
- While source systems are transaction or process-oriented, a data warehouse is subject-oriented to support flexible analytical processing of high volumes of data
- While data stored in source systems is generally real time or near real time, a data warehouse stores historical data
- Data imported into a data warehouse is generally read-only and nonvolatile. Once loaded, it rarely changes – thus preserving an ever-growing history of information

While data can be pulled from source systems, it's not ideal. If customer analytics is a priority, it's worth exploring the options available for starting a simple data warehouse.

Useful Resources

- Rensselaer Data Warehouse Project (Rensselaer)

THE PROBLEM WITH PULLING DATA DIRECTLY FROM SOURCE SYSTEMS

While data can be pulled out of source systems for analysis, that's not ideal for a number of reasons:

- Source systems generally only show the current situation and historical data is often overwritten. This means that while a current snapshot can be seen, behavioral change over time cannot be assessed
- Pulling data out of a source system and preparing it for analysis may be difficult and may compromise stability of the source system

CURRENTLY THERE IS NO DATA WAREHOUSE. WHAT NOW?

- It's worth exploring the investment required to develop a data warehouse. Process and investment depends on a number of factors and requires further research on the part of your organization
- Developing a full data warehouse may not be viable at this time due to lack of capacity to set it up, high costs, and/or financial constraints
- Interim solutions include taking periodic snapshots of data held in your source systems and building a database over time
- Start by identifying what you want to analyze and which source system holds that data
- This will provide some history that may be useful for analysis

Useful Resources

- Data Warehousing for Business Intelligence *(Coursera)*
- Setting Up and Managing a Data Warehouse (Cleverism)
- 6 data warehouse design mistakes to avoid *(ComputerWeekly)*

PART B

Do we have a **single customer view?**

Having a ***single customer view (SCV)*** is imperative in helping you fully understand your customers.

BEFORE
FRAGMENTED VIEW OF CUSTOMERS

WHAT'S A SINGLE CUSTOMER VIEW?

Organizations often only have a fragmented view of their customers. For example, customer application data and transactional behavioral data may be looked at independently, but the same customer may not be identifiable in both data sets. Likewise, an organization may not be able to identify customers who use multiple products or several versions of the same product.

- A single customer view links various data sources via a unique identifier that matches and brings together all data on an individual customer. This single view creates a holistic picture of every interaction with that customer and the services they use

WHY IS A SINGLE CUSTOMER VIEW IMPORTANT?

- It allows your organization to understand how customers behave across multiple touchpoints and understand their cross-product holdings
- It allows you to understand a customer's history, lifetime value, risk, potential exposure to debt, propensity to buy new goods and services, and the number of products they hold

Useful Resources

- Exploiting the Single Customer View to Maximise the Value of Customer Relationships (Experian)

How to Achieve a **Single Customer View**

CREATING A SINGLE CUSTOMER VIEW

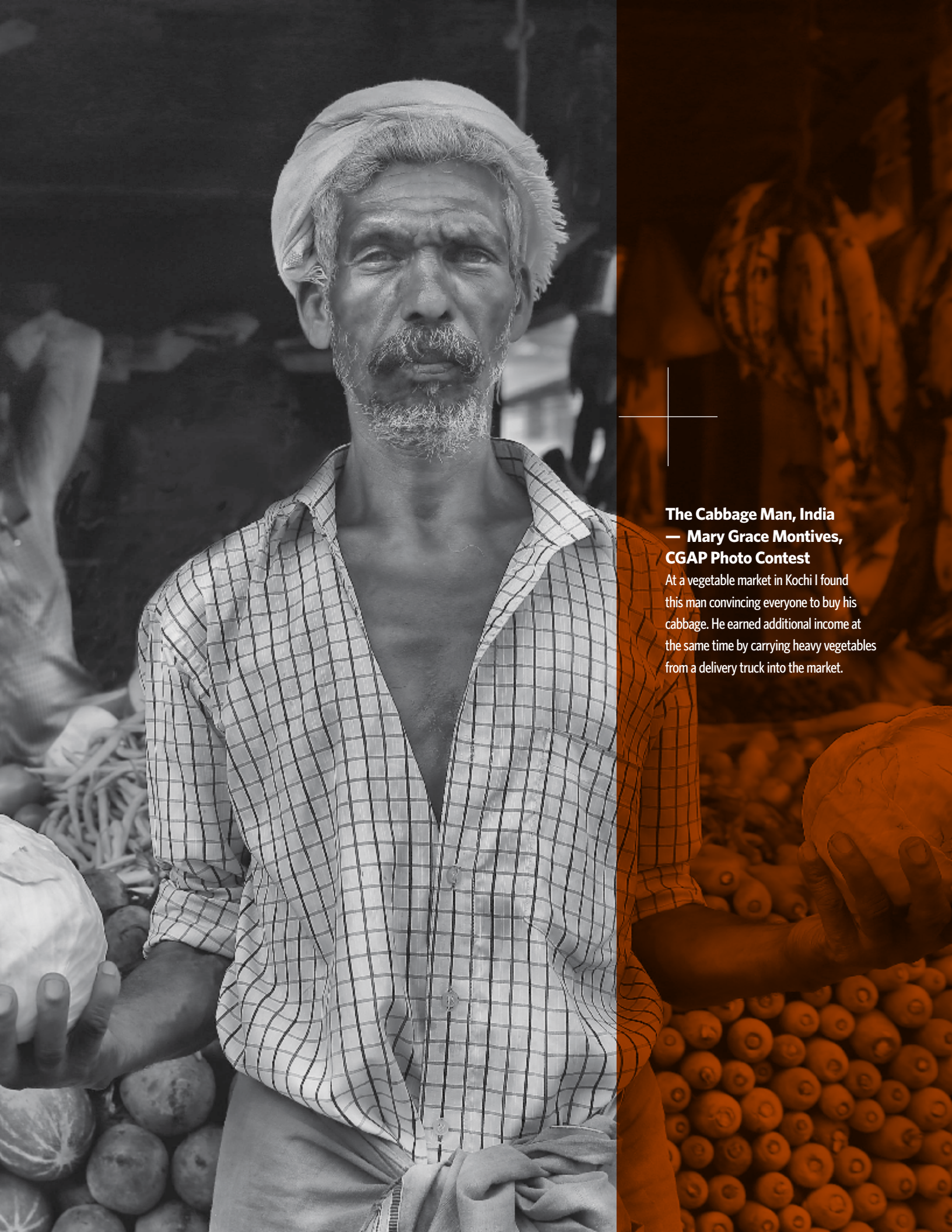
- Requires a customer's product relationships to be brought together
- Can be done by matching unique identifiers, which may be a national ID number (where available)
- If a national ID is not available, other identifiers need to be used such as matching name, address, and date of birth
- Once individual customers have been identified, a unique customer identifier (often described as a customer PIN) should be applied
- The identifier is then appended to every item of data relating to that customer, allowing all data on an individual customer to be matched and brought together to create a single view or record of that customer
- All future data linked to that customer must include their unique identifier (i.e., when the customer purchases a new product or service)

POTENTIAL BARRIERS TO CREATING A SINGLE CUSTOMER VIEW

- The major obstacle is customer data that's stored on different systems in inconsistent formats
- This is a greater problem for large, more complex organizations such as banks with various product streams, or organizations that have undergone mergers and acquisitions which may have multiple brands under a group umbrella
- Product-centric organizations will also find this more challenging, as adopting a single customer view requires a customer-centric approach to customer management. For example, banks where home loans are kept separate from retail banking, etc.
- Smaller or newer organizations with newer technology and fewer product relationships find it easier to implement a single customer view

Useful Resources

- How to Achieve the Single Customer View (Experian)



The Cabbage Man, India — Mary Grace Montives, CGAP Photo Contest

At a vegetable market in Kochi I found this man convincing everyone to buy his cabbage. He earned additional income at the same time by carrying heavy vegetables from a delivery truck into the market.

Benefits of a **Single Customer View**

CUSTOMER BENEFITS

- Enhanced customer service levels: customers expect providers to have a complete view of their relationships and a single customer view makes it a reality
- Reduction in duplicate communications
- Competitively priced products based on the overall customer relationship, e.g., customers with significant savings may be offered discounted personal loan rates
- Data changes are simplified, for example, a change of address notification only requires a single phone call
- Cross-sell and up-sell marketing offers are based on the whole customer relationship and are more relevant

BUSINESS BENEFITS

- Customer service agents see a customer's complete product holdings and history at a glance, which enables them to quickly assess the relationship and take appropriate action
- Operational cost savings as a result of reductions in customer call times and cost per serve
- More accurately targeted marketing offers and incentives
- Better customer understanding drives more effective product development and ensures that the right product propositions are delivered to the right customers at the right time
- Valuable customer insights allow more effective customer-level marketing and improve relationships, cross-sell activities, product penetration, and retention

Useful Resources

- Exploiting the Single Customer View to Maximise the Value of Customer Relationships (Experian)

Thinking Creatively **about Data Collection**

You may operate in a "low-touch" environment with limited data being collected from customers. If so, think creatively about data collection.

PARTNERSHIPS

- Partnerships with organizations that collect large amounts of data may be hugely beneficial
- For example, a partnership with a mobile network operator can provide access to numerous data points, including expenditure on airtime, location, engagement with mobile wallets and accounts, number of contacts
- Partnerships can also complicate the data collection process. Multiple considerations should be reviewed when a partnership is initially formed, including who owns the data and who may access it
- Linking data to internal data sources may be tricky

DEVICES

- Embedding electronics, software, sensors, and network connectivity into everyday objects to enable them to collect and exchange data
- For example, M-KOPA, a Kenyan solar energy company that sells home solar systems in Kenya, Tanzania, and Uganda, generates data from sensors in their solar panels which is then used in loan decision-making (see **Source** below)
- The use of telematics devices in vehicles by insurance providers is another example

LOYALTY PROGRAMS

- Low-touch organizations can increase data collection through loyalty programs
- This strategy is often employed by retailers whose data does not identify specific customers
- Data can be collected from enrollment in and customer engagement with the program
- Organizations need to consider the costs associated with setting up and maintaining loyalty programs

Source

- Now you see me: How alternative data is unlocking new markets for financial services

Using Alternative **Data Sources**

This toolkit focuses on the use of internal data. However, there are numerous external data sources you can use to better understand your customers.

It's useful to look at alternative data sources when:

- You have customer data but there are gaps in understanding, for example, their living conditions, usage of media, access to various service points, disposable income, etc.
- You operate a low-touch organization and collect limited customer data
- You have customer data but are unsure of its quality and would like to validate it using external data sources

Examples of **Alternative Data Sources**

Alternative data sources can validate or augment internal data and help you more fully understand the lives of customers.

EASE OF ACCESS

EASIER

MORE DIFFICULT

SECONDARY DATA

- Surveys and qualitative data generated by statistical and other organizations

DATA FROM VARIOUS ASSOCIATIONS

- Associations often collect and publish aggregate data for an industry

CREDIT BUREAU DATA

- In countries with credit bureaus, a snapshot may be available of account status, balance, contractual status, etc.

PAYMENTS DATA

- Data that can be pulled from external payment systems

SOCIAL MEDIA DATA

- Sentiment data
- Number of complaints, queries, etc.

CLAIMS DATA

- Aggregate claims data from across the industry is frequently published

BILL PAYMENTS DATA

- Aggregate data on bill payments

TELECOM DATA

- Airtime purchases, for example. Top-up and utilization data can provide insights into disposable income and expenditure patterns

Innovative Ways Financial Service Providers **Use Alternative Data**

EXAMPLE 1

In Kenya, the **Commercial Bank of Africa** (CBA) and **Safaricom** partnered to offer credit services (M-Shwari) to 20 million M-Pesa users. M-Pesa began as a mobile money transfer service in 2007. M-Shwari uses M-Pesa usage plus Safaricom historical mobile phone data to assess creditworthiness and offer customer loans. Data including airtime purchases and credit, M-Pesa balance, M-Shwari savings activities, call records, data usage, and customer lifetime length is used to construct a credit score and credit limit. To date, CBA has disbursed more than 20 million loans to almost 3 million unique borrowers and reports a 90-day non-performing rate of only 2 percent on M-Shwari loans (Cook & Mckay, 2015).

EXAMPLE 2

Branch, a digital credit provider, scrapes an applicant's mobile phone for thousands of data points, including contact list, phone make, model, and GPS location when a customer downloads their mobile app and applies for a loan. Through machine learning algorithms, Branch assesses creditworthiness and inputs the loan into the applicant's M-Pesa account in minutes.

EXAMPLE 3

Segovia, an innovative bulk payment provider, uses satellite imagery and biometrics – coupled with imaging analytics and machine learning – to better deliver payments to recipients of development or aid programs.

EXAMPLE 4

Social Lender is an up-and-coming lending solution based on an applicant's reputation on social media platforms. Currently operational in Nigeria in partnership with Sterling Bank, Social Lender develops social reputation scores using social media data and a proprietary algorithm. Loans are guaranteed by the user's social profile and network, which then allows them to borrow funds based on their social reputation rather than a traditional credit score. Users can access their money via existing banking channels and the company is exploring how to disburse credits through mobile money.

Source

- Now you see me: How alternative data is unlocking new markets for financial services (Insight2Impact)

Examples of **Qualitative Research**

Quantitative analysis can test specific hypotheses and its statistical nature allows for generalization. However, it does not offer the level of detail that can be gathered from qualitative data.

Using a combination of qualitative and quantitative data can improve an evaluation by ensuring that the limitations of one data type are balanced by the strengths of another.

FOCUS GROUPS

- Involve hosting a discussion group, usually 6-10 people
- Participants may be current or potential customers
- A moderator guides conversation and ensures that all participants are given an opportunity to contribute

INTERVIEWS

- May be short, targeted intercept interviews conducted with customers or potential customers in their place of business or living
- May be longer, in-depth discussions that probe experiences, motivations, and perceptions

MYSTERY SHOPPING

- A useful technique for better understanding the experience of a customer at various touchpoints
- Particularly useful for testing the quality of service that customers experience

OBSERVING CUSTOMERS

- Many insights can be gained by spending time on the floor of a branch office and watching the way customers act or how they interact with staff

Useful Resources

- CGAP Customer Experience Toolkit

The Chicken Vendor of Kathmandu, Nepal — Anindya Majumdar, CGAP Photo Contest

This vendor comes from his village with a basket full of chickens – the sole source of his income. He seems happy despite the low income and financial instability of his job.

Qualitative Data Can Be Used to **Support Quantitative Analysis**

Small or qualitative data can be extremely helpful in gathering deeper insights around customer experience, preferences, and pain points.

DIRECT QUANTITATIVE ANALYSIS

- Generate hypotheses through qualitative work to be tested through the quantitative approach
- For example, quantitative analysis can monitor discussions with customers and identify key pain points along the customer journey

ENRICH QUANTITATIVE ANALYSIS

- Qualitative research can be used to identify issues or obtain customer information not currently captured by internal data
- For example, customer circumstances, motivations, preferences, and attitudes are not captured well by quantitative analysis

EXPLAIN FINDINGS FROM QUANTITATIVE ANALYSIS

- Using qualitative data to understand unanticipated results or trends
- For example, while transaction behavior can be identified from quantitative analysis, uncovering the underlying reasons for it may require qualitative research

VALIDATE FINDINGS

- Qualitative data can be used to verify results from quantitative analysis (and vice versa)
- For example, where quantitative analysis indicates that certain servicing channels are not utilized, mystery shopping can test them to ascertain accessibility and ease of use

Important Considerations When Assessing **Possibilities Given the Available Data**

THINK CREATIVELY ABOUT POTENTIAL DATA SOURCES

- Consider all potentially useful data sources – not just obvious ones

OPTIMIZE DATA COLLECTION

- During the process of understanding which data is collected, stored, and where, think about how to optimize the process
- For example, if customers are asked for unnecessary data, can that data point be dropped or replaced with one that's more useful? Is certain data collected on application forms but not captured or stored?

CREATE AND MAINTAIN A DATA DICTIONARY

- Ensure that you create a data dictionary and that it's maintained. Include data meaning, relationship to other data, origin, usage, and format
- Identify a data team member to maintain and update the data dictionary

Phase 3
Refine Plan, Data Request, and Analysis

Phase 1
Visioning

Part A
Who's going to lead the initiative?

Part B
What do we want to know about our customers?

Phase 2
Assess Possibilities

Part A
What's possible given the data?

Part B
Do we have a single customer view?

Phase 3
Refine Plan, Data Request, and Analysis

Part A
How do we obtain and prepare data for analysis?

Part B
How can we effectively conduct the analysis?

Phase 4
Dissemination and Engagement

Part A
What kinds of outputs work best?

Part B
Return to Phase 1: Feed back into the ongoing visioning process

PART A

How do we obtain and **prepare data for analysis?**

Now that you've established which data is actually available, prepare to undertake analysis.

Four steps are suggested when preparing for actual analysis:

1

DEVELOP A PROJECT PLAN

- Team lead refines the vision (if necessary), develops a project plan, and identifies project team members

2

SUBMIT A DATA REQUEST

- Analysts submit a data request to the data team, based on the data required for the analysis
- Data team extracts the requested data for the analysts

3

GAIN A DETAILED UNDERSTANDING OF THE DATA

- Analysts spend time analyzing the data structure and data quality

4

VALIDATE THE DATA

- Analysts validate the data to ensure that the data request was performed correctly

Vendedora de Papas, Peru — Carloman Macidiano Céspedes Riojas, CGAP Photo Contest

One form of labor inclusion for women in the Peruvian Andes is the sale of agricultural products that they plant and harvest themselves.

Step 1: Develop a **Project Plan**

A project plan defines the proposed scope, timelines, and the specific team that will be involved in the project. The plan is important for managing expectations and ensuring that everyone starts at the same place.

IMPORTANCE OF A PLAN

- There's now considerably more information about which data is available and what analysis is possible than during the initial visioning phase. At this point, it's possible to create a more detailed project plan
- The plan ensures that the project team knows what's expected and manages expectations of project stakeholders
- Remember that the plan is not set in stone since closer inspection may reveal that certain analyses are not possible or worthwhile given the available data. However, it's imperative that there's an understanding of which customer insights are required to provide the analytics team with direction

WHAT TO INCLUDE IN THE PLAN

- A description of the proposed analysis to be run
- An overview of proposed outputs (deliverables)
- Timelines for completing the project, broken out by task where possible
- The project team that will work on the project. In addition to the team lead identified at the beginning of the process, includes members of the data team and internal or external analysts

Resources Required for Analytics: **Data and Analytics Teams**

Resources required to support the customer insights champion will also be determined in Phase 3.

WHO IS THE DATA TEAM?

- May include one or more members of the IT team, database administrators, and/or developers

DATA TEAM: SKILLS REQUIRED

The data team must have the technical knowledge and skills required to:

- Manage multiple data sources
- Extract data required for analysis
- Be familiar with the data, its quality and reliability, and ensure this information is communicated to the analytics team

DATA TEAM KEY RESPONSIBILITY AREAS

- Coordinate data extraction, ensure that requirements are met, and check/validate data before sending it to analysts
- Source internal reports and/or other data that can be used to validate the data (management accounts, business banking data, etc.)

NUMBER OF ANALYSTS REQUIRED

- The analytics team may consist of one or more analysts, depending on the amount of analytics required
- If your organization is just starting to look at internal data, start with one analyst and grow from there

ANALYSTS: SKILLS REQUIRED

- The technical skills to analyze large quantities of data and the soft skills to identify correct questions, interpret results, and generate engaging outputs
- Technical skills should include a knowledge of database languages (e.g., SQL), knowledge of at least one statistical language (e.g., R/SAS), and Excel skills
- Programming and data visualization skills are beneficial

ANALYTICS TEAM KEY RESPONSIBILITY AREAS

- Assesses what's possible given available data, runs the actual analysis, interprets results, and develops and communicates outputs back to the organization

INTERNAL VS. EXTERNAL ANALYSTS

- Can be internal or external and brought in on a project-by-project basis
- With external analysts, ensure that regular meetings are scheduled with the data team and customer insights champion to keep everyone on the same page
- The decision to use internal or external analysts depends on how much work is required. Often it doesn't make sense to hire a full-time data analyst when an external analytics company can fill the gap

Step 2: Submit a **Data Request**

When beginning to conduct customer analytics, analysts will likely need to submit a request for the data team to pull specific data. Organizations further along the journey may have an interface that allows analysts to pull data themselves. (If that's the case, skip this step.)

Project analysts submit a request to the data team based on their needs for the specific analysis. Here's the process for submitting a data request:

STATE THE OBJECTIVE

- Provide a brief description of the objective of the data extract

COMPILE DETAILED DATA REQUIREMENTS

- Compile into a single document
- A shared document such as a Google sheet is useful

1. It's worthwhile to estimate the total size of the data extract
2. If the data extract is large (over 1 million records, for example), first request a data sample to minimize time required for the initial data extract. If a number of tables interact, pick a sample of customers and request all data related to them (to ensure you can see the linkages between data tables)
3. The sample enables analysts to gain a better understanding of the data and make required adjustments before submitting a full request

A Sample **Data Request**

Group Policy File

Field Name	Description	Notes
GROUPPOL_MST_SEQNO	Unique row ID	
POLNO	Group Policy Number	Used for validation
COVEFF_DATE	Coverage Effective Date	
STRDATE	Enrollment Acceptance Start Date	
ENDDATE	Enrollment Acceptance Start Date	
TUGTAG	Top Up Tag	
DEFPAYTAG	Deferred Payment Tag	
DISTCHAN	Distribution Channel	

Filter on
(STRDATE < 1 October 2016 OR STRDATE IS NULL) AND
(ENDDATE > 31 December 2013 OR ENDDATE IS NULL) AND
POLNO IN codes applicable to CPMI microinsurance products

Member File

Field Name	Description	Notes
MEMBERID	Unique Row ID	
DOC	Date of Birth	
GENDER	Gender	
CIVILSTAT	Civil Status	
ADDR	Address	We will need to see if there is an easy way to parse this field to obtain location
OCCUPATN	Occupation	
REF_MEMBERID	Principal Member ID	

Filter on
MEMBERID IN filtered Enrollment File

Group Policy File

Field Name	Description	Notes
GROUPPOL_PLAN_SEQNO	Unique Row ID	
GROUPPOL_MST_SEQNO	Foreign Key in Group Policy File	
POLNO	Plan Policy Number	
PLANGRP	Plan Specific Name	
RELTYPE	Relationship Type	Used for validation
PLANCODE	Coverage Plan Code	
FACEAMT	Insured Amount	
PREMIUM	Premium Amount	
AGEFROM	Valid Maximum Age	Used for validation
AGETO	Valid Maximum Age	Used for validation
SHORTDESC	Description of Coverage Plan	Used for validation
GRPPOLNO	Group Policy Number (POLNO in Group Policy File)	Used for validation
TERM	Term of Coverage (months)	Used for validation
TERMTAG	Fixed Term Tag	Used for validation
NUMDAY	Claim Number of Days	Used for validation
PREMADJUST	Partned Premium Adjustment	
STRDATE	Enrollment Acceptance Start Date	Used for validation
ENDDATE	Enrollment Acceptance Start Date	Used for validation

Filter on
(STRDATE < 1 October 2016 OR STRDATE IS NULL) AND
(ENDDATE > 31 December 2013 OR ENDDATE IS NULL) AND
POLNO IN codes applicable to CPMI microinsurance products

The data request process can be complex. This is especially true for larger organizations with multiple divisions and a centralized data team, such as:

Case Study: The Data Request Process

During an analytics project aimed at redesigning the reporting process for a division of a large insurer, the data team (made up of a central IT department) controlled the data source. Accessing data was a lengthy, formal process with long turnaround times. Key issues that created delays were:

- Limited capacity within the data team to work on requests
- A data team that was far removed from the project and lacked the context required for requests – which lead to numerous errors
- An analytics team without direct access to the data team, so communication was slow and often misunderstood or misinterpreted
- The division the analytics team worked with had requested a simple database for direct access to required data. The request was repeatedly denied since the organization was creating a long-term solution for all data problems. However, this "super project" was running two to three years behind schedule and had no set end date

HOW COULD THIS HAVE BEEN IMPROVED?

A number of interventions would have vastly improved turnaround times on the project, including:

- Allowing the analytics team and the data team to communicate directly
- Including members of the data team in the project from the beginning so they'd have the required context to help with the data extract
- Allowing divisions to create simple databases without a long, formal process

A similar situation occurred when conducting analytics for a large bank:

Best Practice: Optimizing Data Access

Established banks often have siloed divisions and legacy systems that make it difficult to have a clear, holistic view of the customer data they hold. As a result, both quality and turnaround time on data requests are reduced.

Analysts often have to request the same data set multiple times before they get the correct data. Since they're required to use "formal" channels rather than requesting data directly from the bank's data warehouse or business units, problems are often exacerbated. This creates miscommunication between the analysts (the data users) and the data team (the data providers).

Organizations that want to make customer analytics a priority need to ensure that analysts have easy access to the data they require.

Balance, Bangladesh — Mauro De Bettio, CGAP Photo Contest

As a fisherman balances his daily catch in a large basket on his head, he delicately glances at the camera. His smile reflects tenderness and candor, and this portrait evokes a time and place where things run slower.

Data Security **and Privacy**

When extracting and transferring data, the data team must be aware of security and privacy issues.

- Organizations, especially financial service providers, hold a considerable amount of customer information, including personal details and other sensitive data
- Any processing of this data threatens customer privacy
- A number of countries have legislation around protecting personal information. But even where this does not exist, it's important to take appropriate precautions to avoid reputational risk through the loss or misuse of personal data
- Implementing a data standard can help control data sharing
- A data standard specifies responsibility, accountability, policies, and procedures on data sharing and analytics. This ensures that the correct person signs off on data extracts, and that shared data is "safe" (i.e., does not include unnecessary sensitive information) and safely transferred
- Hashing algorithms can be used to mask personal identifiers or depersonalize data. Common hash functions include MD5 and SHA1. (Hashing is used to enable database items to be more quickly retrieved and in encryption/decryption of digital signatures. The hash function transforms the digital signature, then both the hash value and signature are sent to the receiver. The receiver uses the same hash function to generate the hash value and then compares it to that received with the message. If hash values are the same, it's likely that the message was transmitted without errors. These details are particularly important when data is shared with external analysts)

Source

- What does Hash Function mean? (Technopedia)

IMPLEMENT A DATA STANDARD

- Implement a data standard that specifies a process to be followed when data is extracted and shared

Useful Resources

- Ten tips to protect your customers' personal information (OAIC.gov)
- Building consumer trust: Protecting personal data in the consumer product industry (Deloitte University Press)

Transferring **Data**

When an analytics team is external to your organization, it's critical to securely transfer data to them.

- When data needs to be transferred, a **Secure File Transfer Protocol (SFTP)** is the most efficient solution
- Once data is transferred, the external analytics team must check the integrity to ensure data hasn't been corrupted in the process
- The analytics team can conduct simple checks, such as confirming row and column count
- Hash-based verification can also be used to check data integrity. The data team uses a hashing algorithm to create a hash and sends it, along with the data file, to the analytics team. The analytics team uses the same algorithm on the data file to ensure that the resulting hash matches the data team's hash
- When an SFTP cannot be set up, an alternative method is to physically transfer data on an encrypted hard drive

Even when a data standard is in place, breaches may occur:

Case Study: A Potential Data Breach

During an analytics project, a bank and an external analytics team spent months organizing data privacy protocols. This included hashing all sensitive data fields to protect account holder identities and account details.

When the data extract was eventually passed along to the external analytics team, data included personal information that should have been hashed. The data team member who pulled the data did not correctly apply the necessary adjustments before passing it on, and there was no process to check the extract before it was sent.

While the issue was immediately flagged and rectified, if the data had been used inappropriately the breach would have resulted in a large reputational risk to the bank and the external analysts.

HOW CAN A BREACH BE AVOIDED?

- It's imperative that all members of a data team are aware of data standards and understand which data is sensitive, cannot be shared, or requires masking
- A senior member of the data team should sign off on any requests that may contain sensitive information

Step 3: Gain a Detailed Understanding **of the Data**

Once a data sample has been received, it's critical that analysts spend time understanding the data and assessing its quality.

UNDERSTANDING THE DATA STRUCTURE

- While Phase 2 indicated which data is available, it's only once a data set or data sample is acquired that the analytics team can really understand how it's structured
- It's important for the analytics team to take the time to understand every data field and confirm the meaning of all coded fields
- A data dictionary is critical in helping avoid assumptions. A slight misinterpretation of a field can have major implications on an analysis
- Analysts should ensure that all tables link correctly and as expected, and that business rules are logical

UNDERSTANDING DATA QUALITY

- The analytics team must understand the limitations of the data. Specific fields may be known to be incorrectly captured or not fully captured. Others may be dynamic and expected to have changed since they were captured (income, for example)
- There may be wider issues with the data, such as counting customers twice or including inactive customers
- In some cases analysts can clean data or replace unreliable data with derived fields. A derived income field can be generated based on assumptions about income growth, adjusted by the age and education level of the customer if available. Likewise, address data may be poorly captured and could be replaced with area data collected at channel points, etc.

Encourage communication during Phase 3. **Clear communication between data and analytics teams is imperative**

Joy!, Kenya
— Wim Opmeer,
CGAP Photo Contest

A combination of capital and microloans provided by the Rabobank Foundation – along with Food and Agriculture Organization training – ensures that farmers can buy good seed and hire workers for planting and weeding.

Spending time understanding the data before jumping into the analysis is imperative:

Case Study: Understanding Data Prior to Analysis

On close inspection of customer data prior to an analysis for a large insurer, the analytics team found various anomalies.

Some links between tables were incorrect. Data in the enrollments table is supposed to link to a policy plan file that defines the policy and its cost. However, values included in both the enrollments table and the policy plan table did not align due to misclassifications and lack of standardization in the data input process.

A wider issue identified by the analytics team was incomplete data in the enrollments table. Not all enrollments were captured, and some were only captured when the customer put through a claim. The implication was that the enrollments table was incomplete; if the analytics team were only to analyze this data they would get a biased view since there was a higher chance a customer would appear in the enrollments table if they had made a claim.

The analytics team reported all anomalies to the data team so processes could be developed to improve data quality going forward.

Attempting to make sense of free text fields can be challenging:

Case Study: Free Text Fields and Fuzzy Matching

In an analysis of credit consumers using data from a credit bureau, the analytics team wanted to focus on a specific geographic area. But since the address field was a "free text" field and people spelled areas differently, the data was not useful in its raw form.

Initially the analytics team decided just to pull out postal codes. However, even these codes were not clean enough to reliably identify areas.

The analytics team decided to utilize "fuzzy matching" – a technique that matches similar text strings in instances where exact matches do not work. The team was able to identify and group addresses into different areas and use the newly defined field for analysis.

This same technique can be used with personal identifiers (name, address, etc.) to find individual customers and create a single customer view.

A similar technique was used for analyzing customer data from a large insurance company.

When making a claim on a property, customers were asked to explain the incident type (for example, flood or typhoon). Since this was a free text field, analysis of data in its original form was impossible.

The analytics team used fuzzy matching to classify incidents into a number of predefined options. The team was able to classify a majority of the data, making a previously unworkable data field useful for analysis.

Step 4: Validate **the Data**

Once the full data set is received, it must be validated to ensure that the data ties back to other business metrics.

VALIDATING THE DATA

The analytics team needs to think critically about the data to ensure that it's sensible – and validate data points to ensure they align with other business metrics.

- Ensure that data aligns with expectations – are there the number of rows expected?
- Request a copy of the SQL script (if applicable) used to extract data to ensure there are no obvious errors
- Check for quality issues such as duplicate entries or data that links to incorrect tables. The data team may not be aware of these issues
- Check that business rules work and make sense
- Summarize key metrics and validate against business expectations and independent sources by aggregating data for a specific period and comparing results to independent sources
- For validation, start at a macro level (e.g., number of customers and total value of transactions/balances) then drill down more granularly

IT'S IMPERATIVE TO ALWAYS VALIDATE THE DATA RECEIVED

- The analytics team is often under pressure to go straight into analysis and deliver metrics
- Ensure that adequate time is allocated to properly validate the data, which minimizes errors and saves time in the long term

Important **Considerations**

When the analytics team is external to your organization, data must be transferred to them in a secure manner.

ALLOW ENOUGH TIME FOR THE DATA EXTRACT PROCESS

- The data request and extract process often takes considerably longer than expected. Extraction seems simple but it's generally complex, and includes time to understand the data, check its quality, and validate it. Be sure to allow enough time in project timelines

ENSURE CONTINUITY IN THE DATA TEAM

- Expect a lot of communication between analysts and the data team (IT/database administrators). The data team has large amounts of knowledge about quality and structure that analysts require to ensure that data is used correctly
- It's optimal to use the same data team throughout, although resources often move onto other work before data extraction is complete. If consistency is not possible, have a proper handover process in place and ensure that all processes are adequately documented so new team members can quickly jump in

INVOLVE THE TEAM LEAD TO SUPPORT A CRITICAL APPROACH

- Analysts' questions around data quality and the validation process may be a cause of friction with the data team
- It's important that the team lead supports a critical approach and makes the project a priority for all involved

IMPLEMENT A DATA STANDARD AND MAKE SURE IT'S FOLLOWED

- Specify a process to be followed when data is extracted and shared
- Ensure the data standard is adhered to. When timelines are tight people may look for shortcuts, which may cause data privacy and security issues

PART B

How can we effectively **conduct the analysis?**

It's useful to understand the different types of data analysis that can be performed. There are three broad types:

DESCRIPTIVE ANALYSIS

- Condenses big data into smaller, more useful pieces of information
- Generally involves aggregating and mining data
- Includes reports that summarize the number of customers at a particular point in time – or new customers over the past week or month. May include details such as customer location, products held, preferred servicing channels based on past behavior, etc.

PREDICTIVE ANALYSIS

- Draws insights not directly observable in the data. It's not simply "prediction" as the name may imply
- Includes statistical analysis, such as cluster analysis, to determine like groups, forecasting, and predictive modeling. A common application is using it to produce a credit score

PRESCRIPTIVE ANALYSIS

- Goes beyond predictive analytics by recommending actions and indicating likely outcomes for optimization
- Includes using analytics to optimize pricing, marketing, or servicing channels that improve customer offerings

To draw out behavioral insights, look for interesting patterns and outliers in the analysis.

Useful Resources

- Leveraging Customer Analytics for Business Success (Wharton)
- Big Data Analytics: Descriptive Vs. Predictive Vs. Prescriptive (InformationWeek)

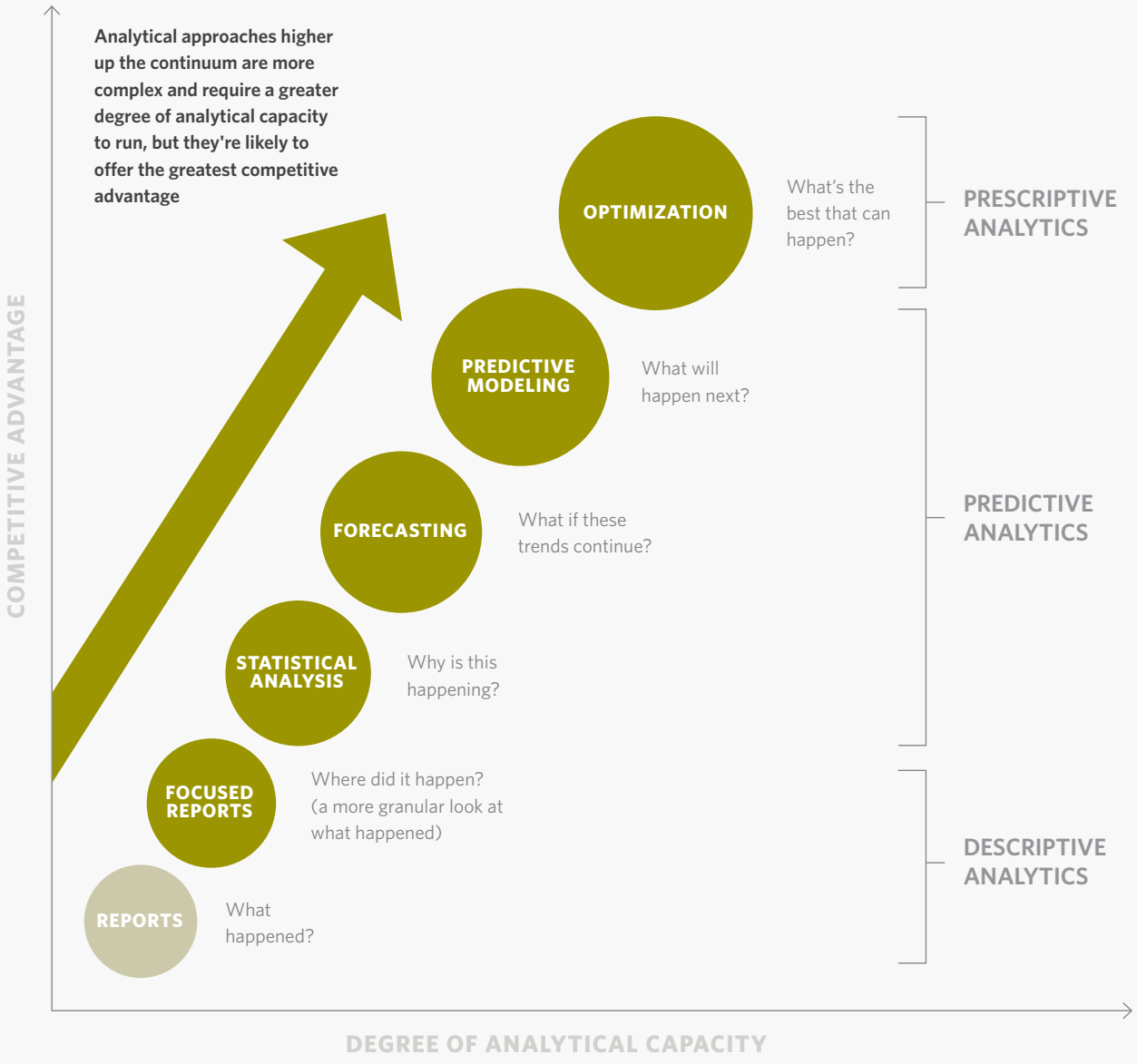
Fish Sorting, India — Abhijit Banerjee, CGAP Photo Contest

Women are engaged in fish sorting at Digha sea beach. Their husbands are fishermen and do sea fishing.

The Woman in the Wind, Vietnam — Thao Vu Xuan, CGAP Photo Contest
A highland woman cleans rice the natural way – in the wind.

Analytical approaches can be displayed on a continuum – **from analytics that produce simple reports to optimization:**

If you're new to customer analytics, start small. Choose an approach that's low cost and easy to implement. Simple analytics can have a high business impact and spark new, more complex questions. As your data and analytical capabilities improve, move along the continuum toward optimization.



Selecting the **Right Tools**

A variety of programming languages and statistical packages can be used to run your analysis.

- Note that the business focus should be on selecting the right analysts rather than the right technology. Analysts need to be able to creatively explore data and pull out insights, while technology is simply a tool to help them do it effectively
- Careful consideration should go into selecting the correct tool – numerous data analysis packages are available and each has its strengths and weaknesses
- For example, a tool may be strong in conducting analysis but weak in building graphics
- Analysts should have knowledge of more than one package so they're able to select the best tool for a specific analysis

Which tool is best largely depends on the type of analysis being conducted. When selecting a particular package, consider the following:

- **Cost** – Is the software open source or commercial? For example, R and Python are open source and can be downloaded for free while SAS is commercial
- **Ease of learning** – Some tools require a steep learning curve while others are more intuitive
- **Graphic capabilities** – Some tools allow users to easily create beautiful data visualizations with charts and dashboards. There are also separate data visualization programs that can be used in combination with data analytics tools
- **Data handling** – Most tools are able to handle large data sets, but some, like Excel, are limited in how much can be analyzed
- **Analysis capability** – Certain tools are more flexible in the types of analysis that can be run
- **Level of support** – Open source programs generally have large online communities but do not have the customer service support associated with commercial programs

Useful Resources

- Python vs. R (vs. SAS) – which tool should I learn? (Analytics Vidhya)
- Data Analytics Tool Comparison: Excel, SPSS and SAS (Angela Ju, Bell)
- How to Pick the Right Statistical Software (The Minitab Blog)

Customer **Segmentation**

When conducting analysis, it's important to remember that not all customers behave the same way. Segmenting customers may reveal otherwise hidden insights.

- Financial service providers have traditionally categorized low-income customers into a single market segment
- Many organizations are beginning to see expanded business potential as they learn that low-income markets are far from just a single segment. People's worries, wants, needs, and behaviors are nuanced, and customer segmentation is the method that explores and identifies opportunities in these differences

BENEFITS OF CONDUCTING A SEGMENTATION

- Improved understanding of current and potential customers
- Addresses business challenges and opportunities
- Identifies and estimates market opportunities
- Helps tailor products, services, and customer experiences
- Helps shape communications

Segmentation can be as simple as using customer demographics to group customers or using a statistical methodology **such as cluster analysis to group customers based on similar behaviors.**

Useful Resources

- CGAP Customer Segmentation Toolkit

Survival for Existence, India — Somraj Sahu, CGAP Photo Contest

On the one hand I see people complaining about everything they have in life and on the other I see this lady who looks fully drenched in flood water, struggling to reach her home to feed her family.

Key Considerations **for the Analysis**

Begin with the end in mind – Although noted in previous phases, this is worth noting again. The analytics team should have an idea of which questions their analysis aims to answer and what the final output will be: a dashboard, a profile of segments, etc. Closer inspection of the data may lead to new questions but the analytics team will need some direction

Start with exploratory data analysis (EDA) – Visually plot the data to get a sense of what it's saying and where you should be looking. This will help the analytics team gain a better sense of the data and what may be interesting in it

Generate reproducible research – It's imperative that the analytics team documents the process – from inputs to final output. Then, if the analysis needs to be run in the future, a clear process document exists

Keep track of data quality issues – Quality issues are often only picked up once the analytics team starts carefully looking at data in the initial stages of analysis and validation. Any issues should be logged with the data team so they can be fixed. The analytics team must always document exactly which adjustments, if any, were made to the data being analyzed

Look for patterns and outliers – To draw behavioral insights from the data, look for patterns and outliers to investigate further. For example, you may see enrollments pick up in a certain month or that a specific branch is experiencing higher-than-average churn. These are indications that you may need to dive deeper into data to better understand trends

Phase 4
Dissemination and Engagement

Phase 1
Visioning

Part A
Who's going to lead the initiative?

Part B
What do we want to know about our customers?

Phase 2
Assess Possibilities

Part A
What's possible given the data?

Part B
Do we have a single customer view?

Phase 3
Refine Plan, Data Request, and Analysis

Part A
How do we obtain and prepare data for analysis?

Part B
How can we effectively conduct the analysis?

Phase 4
Dissemination and Engagement

Part A
What kinds of outputs work best?

Part B
Return to Phase 1: Feed back into the ongoing visioning process

PART A

What kind of outputs **work best?**

Once the analysis has been run, results need to be communicated effectively and disseminated to the rest of the organization.

What Should Outputs Look Like?

There's a wide range of possibile outputs – from static and dynamic dashboards that allow for drill-down to Excel reports, written reports, presentations, etc.

Type of output is influenced by a number of factors, including:

END USER

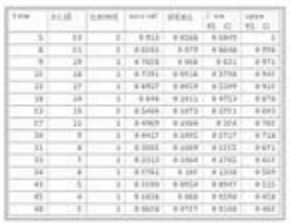
- What's their skill level, what information do they require, do they need to be able to drill down into the data themselves?

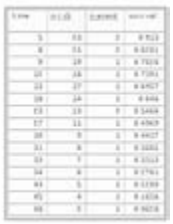
TYPE OF ANALYSIS

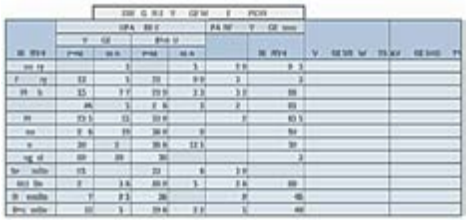
- Is the analysis static or dynamic (changes regularly), can drill-down functionality be implemented, is the output for a specific point in time or trended, is the analysis location-based (i.e., can a map interface be utilized)?
- Ensure that output is meaningful – insights can be lost in aggregated numbers. Allow for drill-down where possible so metrics can be viewed at a granular level

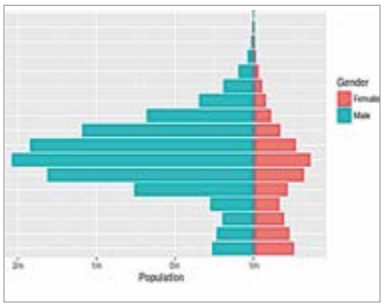
Data **Visualizations**

Analysts need to think about how best to visualize the data.



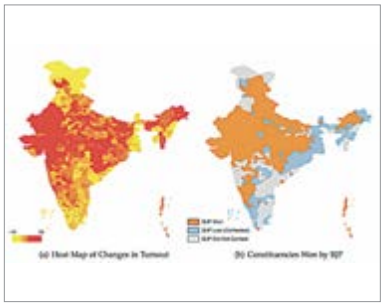


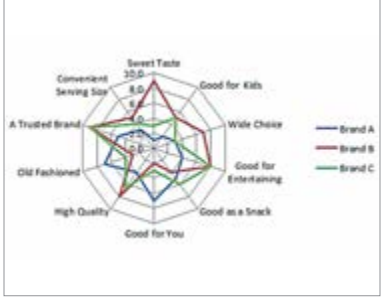


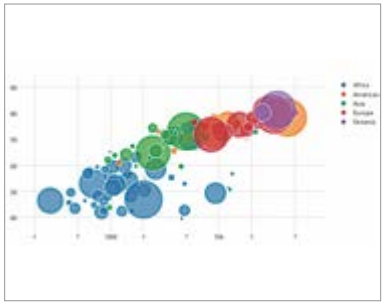


Customer metrics can be graphically displayed to effectively communicate findings

Outputs that link to geographic data are best shown on a map





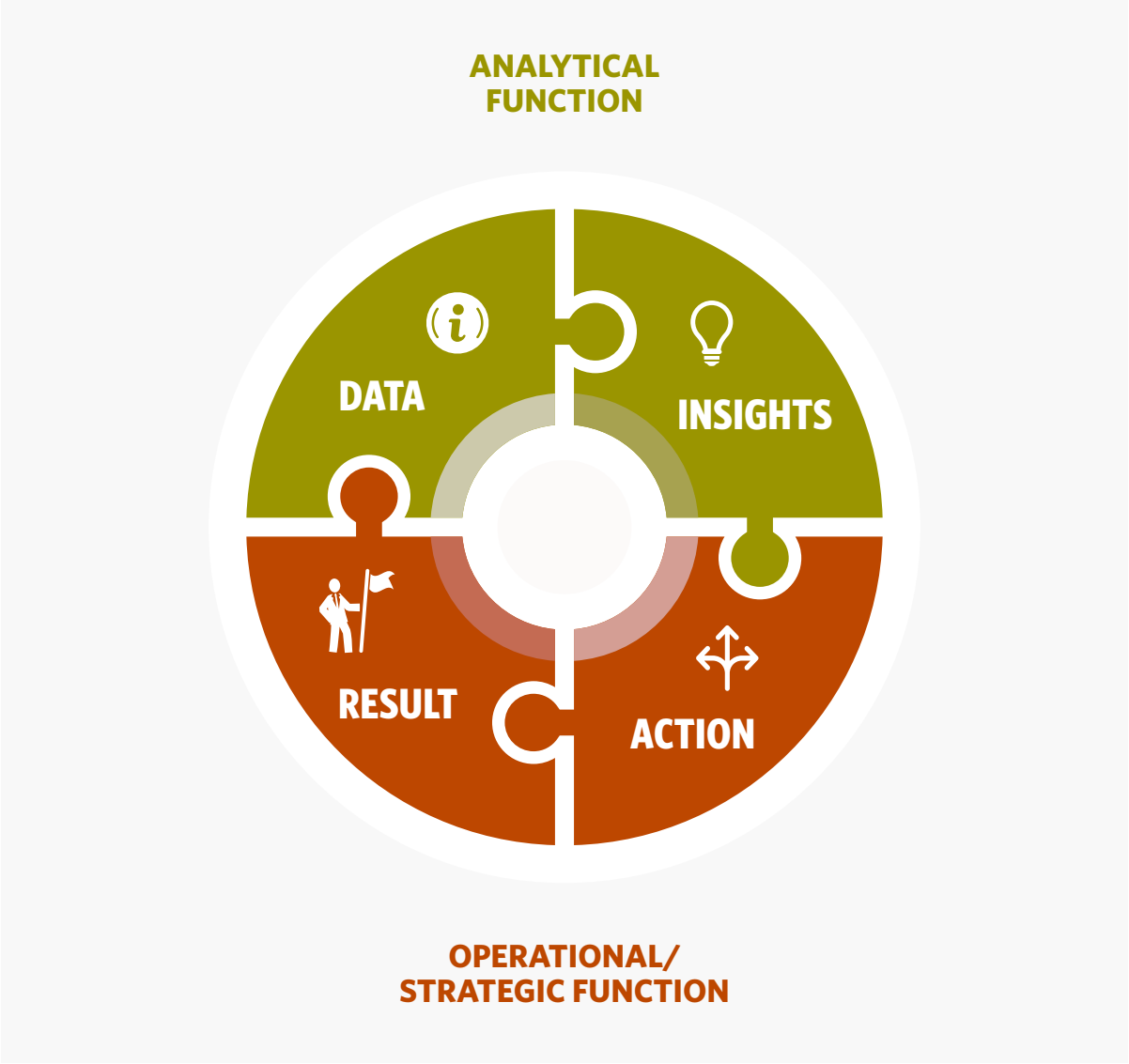


Analytical packages often include various data visualization tools

Red Chili Processing in Bangladesh — Joydeep Mukherjee, CGAP Photo Contest

Women work in a chili pepper drying factory under the hot sun near the Jamuna River in Bogra. Most of the chili comes from river islands and is the main income source for people in the area. They do not have many work options due to climate change.

Although it is easy to dump data into a simple table, there are more engaging ways to display it. Communicating insights effectively is imperative as it encourages engagement with findings. Analysts must creatively consider how to create engaging output.



Insights should be shared across your organization and actions taken based on those insights. Continuously monitor results and make improvements as necessary.

PART B

Return to **Phase 1**

Feed back into the ongoing visioning process.

Undertaking customer analytics is an ongoing process. Through engagement with data, new issues, questions, and avenues for exploration emerge. **Feed them back into your vision and begin the process again.**

Development in Pottery, Bangladesh — Mohammad Reaz Uddin, CGAP Photo Contest

Pottery can be a good source of income for women who become proficient in this art. Women have always produced high quality pottery and currently their contribution to the trade is increasing.

Hard for Life, Vietnam — Thao Vu Xuan, CGAP Photo Contest

Vietnamese fishermen in a little boat on big waves. Is this the fury of the sea over the greed of humans?

Case Study: Card Pioneer Microinsurance Data Analysis Project

South African analytics consultancy Eighty20 worked with Philippines-based Card Pioneer Microinsurance Inc. (CPMI) to assess and analyze the organization's existing data. Their goal was to obtain strong, actionable customer insights. The project's collaborative team followed the process outlined in the CGAP Customer Analytics Toolkit, designed to help financial service providers uncover valuable insights by optimizing internal data which, in turn, improves customer centricity. The following case study describes the engagement.

Phase 1: **Visioning**

First, establish a team that drives the process of generating customer insights from analytics. Initiate a conversation about what the organization seeks to understand about its customers.

At the beginning of the engagement, CPMI established a team that included a lead data analyst to act as the primary contact/project manager, plus additional members from data and IT. The CEO of CPMI was the project's sponsor and internal champion. This leadership ensured buy-in and strengthened the message that generating customer insights was a priority for the organization.

The joint CPMI and Eighty20 project team met early on to identify customer-related focus areas. Given the state of data and systems at CPMI, it was clear that the project would be largely exploratory. The team needed to establish which specific analyses and metrics could even be developed. Potential focus areas were loosely framed around identifiable customer characteristics in existing data: demographic, location-based, and product-holdings indicators. These metrics would eventually be displayed on a user-friendly dashboard.

Sea Man and Sea Fish, Bangladesh
— Rokonuzzaman Khan, CGAP Photo Contest

Dry fish is a seasonal business, and people collect fish from October to May. They dry and process it, then finally sell it in the market.

Phase 2: **Assess the Possibilities**

Next, assess what is possible given the quality, accessibility, and availability of customer data. It is crucial to understand which data is currently collected, how complete and accurate it is, where it is stored, and whether it can be analyzed using a "single customer view."

CPMI had not yet established a data warehouse, so all data needed to be extracted from source systems. The organization was also in the process of creating a single customer view, which would ultimately link individual customers across the full suite of CPMI products they used.

A request was made for an initial sample dataset with a data dictionary and database architecture diagram so the team could review in detail which data existed, along with its quality and completeness.

Several issues were flagged when the dataset came back:

- Free text fields were present.
- Fields were applied in multiple styles across products.
- The sample was not uniform across products, and data was more complete for some products than for others.

The last point reflected different sign-up channels. In some cases, customers signed up as a group so limited detail was captured for individual members. (While group sign-ups may be an efficient practice, there may be implications for the availability and richness of data associated with them.)

Workers, India — Joydeep Mukherjee, CGAP Photo Contest

A microfinance loan helped the owner set up a cast iron workshop that generates employment for local laborers. Catching these working men under brilliant light made a nice composition.

Phase 3: **Refine Plan, Data Request, and Analysis**

During this phase, refine the initial project plan based on available data. Once it is approved, analysts place a data request and run analysis.

After assessing available data, the CPMI/Eighty20 team refined their initial plan to focus on products and customers with associated data that was reasonably complete and reliable.

The data team submitted a detailed data request outlining required fields. (A data request and review process is usually iterative: Analysts receive and review data, request clarifications, and flag inconsistencies. The data team then explores issues, reruns queries, and resubmits data to the analysts.) The CPMI project played out in a similar fashion. Numerous data requests were submitted, and quality and completeness were assessed until the team was satisfied.

Once a final dataset was accepted, the team validated and cleaned it for analysis[1], including:

- Reformatting various fields. Dates, for example, held multiple and sometimes invalid formats.
- Parsing "free text" fields – such as claim type, occupation, and location – to obtain usable field levels.
- Correctly identifying members and products across tables.
- Joining tables to produce an enrollment view and a claim view.

In the field of data analytics, numerous tools exist. For this engagement, data was loaded into an Eighty20 tool that cross-tabulated fields and created visual charts so interesting trends and features would become easily observable. To wrap up this phase, a simple customer segmentation exercise compared enrollments, claims, and other defining characteristics based on gender and age group.

1. The process used Pandas data analytics and manipulation library in Python and loaded into an HP Vertica database.

Trampled Habitat, India — Prabir Talukdar, CGAP Photo Contest

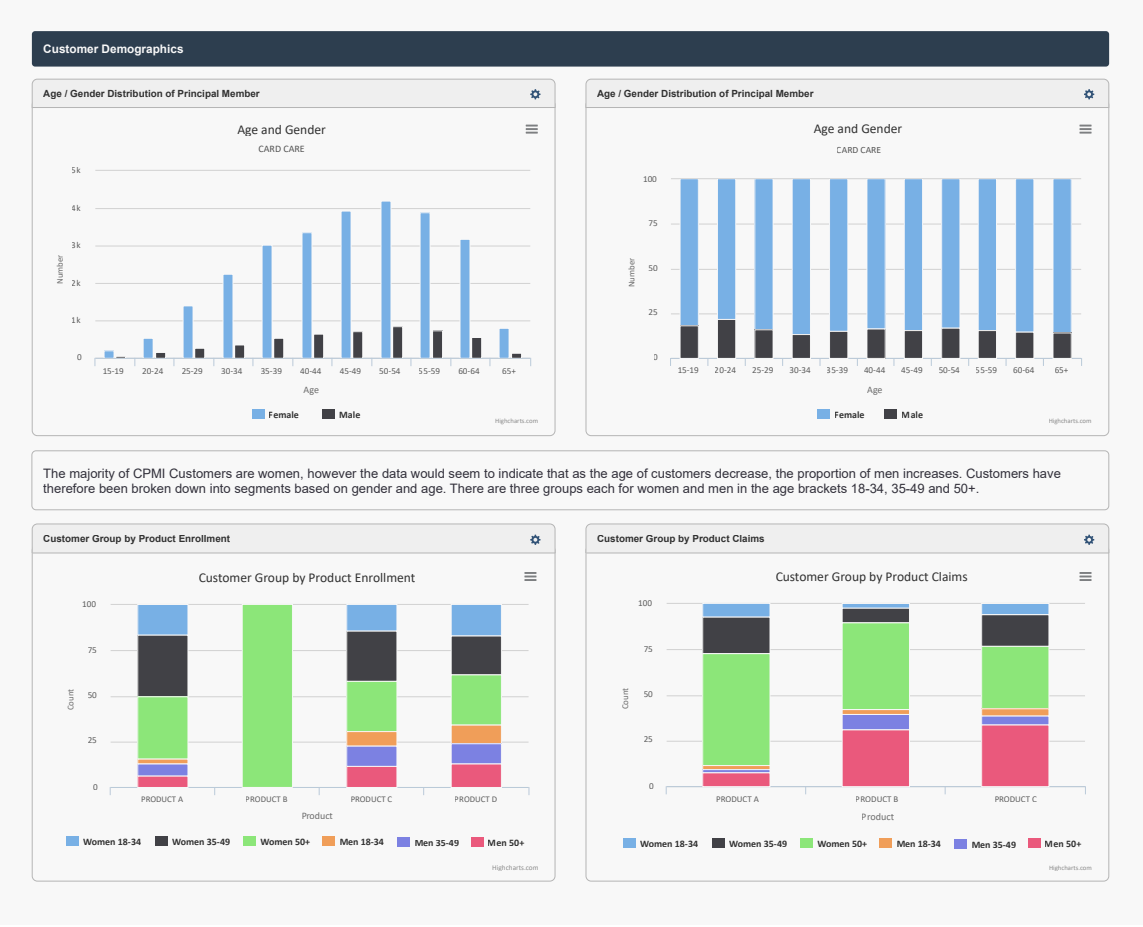
The only fisherwoman of Deepor Beel wades through the polluted water of an area that has been declared wetlands under the 2002 Ramsar Convention. Right now the lake is experiencing conflict between fisherman tribes and industries in the vicinity. Chemicals thrown into the lake have diminished aquatic populations. This woman has been fighting for her rights for a long time.

Phase 4: **Dissemination and Engagement**

During the final phase, effectively communicate and disseminate analysis results to the organization.

A dashboard became this engagement's final output. For members of the management team who only want to engage with headline statistics, it displayed summary metrics. For employees more interested in in-depth exploration of customer data, drill-down functionality included cross tabulation, filtering, and data trending.

There are many other ways to present findings from an analysis. Choice of output is typically a function of the type of analysis being undertaken, as well as analytical capabilities of the user base.



Conclusion

CPMI was at the beginning of their customer-centric journey when the engagement began. Significant team resources were required to generate a sufficiently useable dataset, and the final analysis reflected those limitations.

Three positive outcomes resulted from the project:

- The overall process gave CPMI management an in-depth understanding of the limitations of available data.
- The organization gained a base from which to prioritize mechanisms to enrich data.
- A customer-centric dashboard tangibly demonstrated the potential of further analysis.

Since this engagement, CPMI has worked to improve data quality and systems, and has also developed a customer dashboard that leverages data and provides customers with a tool they are comfortable using.

Eighty20 developed this Toolkit under the guidance of CGAP.
The CGAP team included
Yanina Seltzer, Lisa Stahl, and Gerhard Coetzee.

CGAP

CUSTOMER ANALYTICS TOOLKIT | 2017
DRIVING CUSTOMER CENTRICITY THROUGH ANALYTICS
CGAP.ORG