DATA WAREHOUSING

Project: Data Warehousing Solution for ABC Bank

Deepanjan Rane

1. Company Background:

ABC Bank is one of the top banks in India with over 3 million account holders in 350 branches. ABC Bank has gained success in the past few years experiencing a huge jump in connecting new customers. Not only new customers but it has been successfully working on keeping the existing customers connected with the bank in the long term.

ABC Bank is focused on improving quality and standards and spending more on research and development in the field of banking. They are constantly utilizing resources to deliver better service and products to their customers. ABC bank has products such as credit cards, net banking, mobile banking, savings account, current account, mutual funds, etc. They have been constantly monitoring their performance and keep upgrading as needed.

They were able to achieve this by working on the analytical data that they gathered from bank performance. They closely monitored the movement of resources and strategized their plans accordingly. ABC bank has staff to accept written applications for account opening, loan procedures, insurance policy, etc. which is a hustle work and can lead to human errors.

Today the bank has been connected with such a huge customer chain that they are now working on how to keep the quality and efficiency intact with the bank's growth. Data is one of the crucial parts of the bank. Customer data, data from various products from the bank is one part of bank data, the other major part is the performance data of the bank, which is usually captured from various performance measures and live data from other departments.

Currently, the bank is able to handle customer data and product data, they are using a system that is capable of handling a limited amount of data that is majorly on excel sheets and workbooks. This approach is time-consuming and the system that the bank is using to feed these excel sheets and workbooks is a highly manual task. For ABC bank I am proposing a data warehouse and business intelligence implementation solution to overcome the gaps that they are having in the current system. This proposal will enable them to store their data in a centralized location that can be accessed as per the authorizations and can be efficient and capable enough to process a large number of datasets as required. This will ensure that they are retrieving correct data without and hustle from a single location and with help of business intelligence they can make quick decisions.

This solution will be implemented from the business perspective of the bank where they deal with data from various products such as credit cards, current accounts, saving accounts, online transactions, disputes with various other modules such as ATMs, accounts, etc. Data warehousing will enable ease of access. It will help to retrieve data from various products or departments in form of reports which will help in decision-making and product performance reviews. Collective feedback data from these products will help the bank to work on aspects that are not seen in daily work. The consolidated data help to provide clarity over the working of the current system and help in doing better with the use of business intelligence.

Business intelligence will help in better decision-making for the current system and will also work on making future decisions and strategies to keep the bank growing in its sector. Hopefully, this will fill up the gaps that the bank is currently lacking behind with its growth in the banking sector. This will open doors for new technologies to be introduced in the banks by switching to automating the

processes to have more data scraping with less human effort, this will also help in maintaining data consistency for overall processes. The bank will be integrating its current data handling system into data warehousing with an aim to maintain data integrity and consistency. In ABC bank data warehousing and business intelligence play a vital role in bringing up the business growth and its position in the market within its competitors. The DW/BI is used to make better decisions for the organization. When you are in a market with competitors and own products that are game-changing in your sector you definitely need to maintain that position because when you provide a large number of services and products to your clients and customers you need to keep track of their performance. Customers or clients may be using all of your products or services or only just one product or service and if you fail to deliver it with that quality you will lose your client customer.

So while producing best-in-sector services and products there is a need for constant monitoring and analysis of the performance and working of the products and services, which will ensure intact quality and consistency of the services. This requires consolidated and fast data services which are gained by data warehousing and business intelligence. The bank is looking forward to data-driven solutions that will help them fill in the gaps in the organization and do more progress. This shall be implemented step by step and product by product to ensure seamless integration into the new system. They can work on the current process to take feedback from different departments and note down what more can be added to collect data from that particular department or product which will help them to make effective decisions.

The problems are majorly related to a lack of data for performance evaluation since less data is collected as feedback the authorities don't have a clear view of the products or systems to make further decisions. The analyst or the decision-making authorities aren't known of the live performance data of the bank, which restricts them from not only making future decisions but also current decisions that may cause a lot of issues to the system. That is why it is necessary to have relevant data that can help out to make firm decisions. In this case, the relevant data is stored at one single location can which will help to retrieve data in various mix-and-match cases which aren't specified earlier. The data from the warehouse can be retrieved as per user requirements, for example, details of accounts that withdraw cash from different branches in different locations so that the bank will get to know its account holders about their movement and can suggest the customer better approach. Or the bank can get an overview of car loan product that does not perform unless there is any festive offers or festivals going around in that particular region or state, which can help authorities to make decisions to develop more encouraging loan offers that can lead to long-term customer commitment.

When you have insufficient data it is really tough to make decisions. That's what I am trying to solve in the project by collecting data processing it and making it ready to know more insights into data and taking decisions accordingly. This will not only help in keeping us ahead of the competitors but also help in recognizing our faults and defects in the current system. This will open up areas for growth by knowing more about the performance. And will also allow us to make crucial decisions like adaptation to the market trends.

This deliverable is able to help the bank with making better decisions on products that they have created for customers. They will be able to track the real-time progress of their product which can help them more in understanding the behavior of the product and services. This will also help them forecast the development or updating of particular products that can lead them to success. The data that they will be processing will also lead them to make decisions for updating with new technology or the needs and demands in the market. Understanding these needs and demands from the market and the analytical data available to them they can build or develop more products that can yield more profits or create a platform to become a game changer in its sector. More important decisions such as putting a hold on a particular product or service decisions can be taken from this analytical data.

2. Business Objectives and Imperatives:

ABC Bank is more focused on holding its current customers along with planning to grow in all directions. To do so they need to be ahead of their competitors. To do so they have to keep upgrading and adapting to what's new in the market, not only be ahead of the competitors but also to engage new customers with the bank keeping their current customers happy and keeping the organization in profits. Currently, the bank is not well set up to analyze the data in such a huge quantity. To make it possible new data warehouse is being implemented.

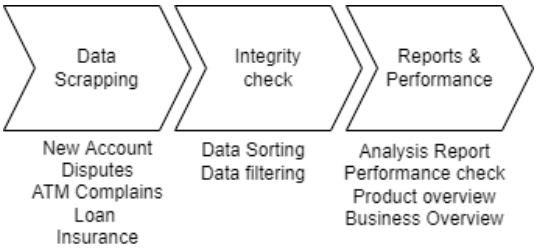
3. Problems and gaps addressed by DW/BI:

The system currently lags in data integrity and consistency. Data from various departments is not stored at one location rather it is distributed at different locations. Generating reports and an overview of the current business is impossible collaboratively. Implementation of this system will ensure that correct data is entering the system and a well-managed log is maintained for data tracking and analyzing the data.

Overall this system will help in overcoming the gaps in the current systems such as:

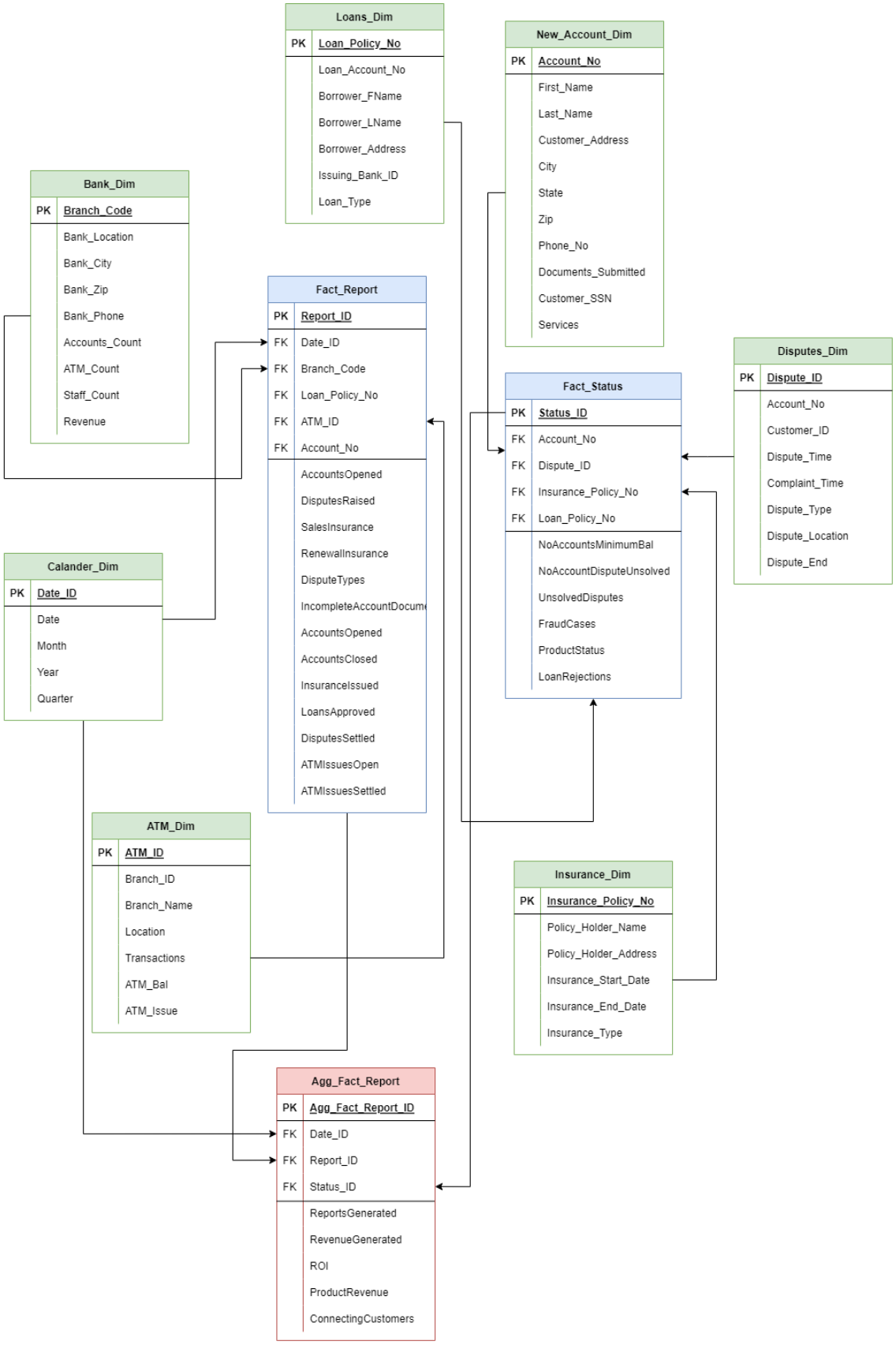
- Inconsistency in data.
- Consolidate data management.
- Data impurities.
- Data-driven analysis of the system.
- Improve decision-making.
- Products overview.

4. Process Maps:



Business Process	KPI/Report/Requirement	Fact	Dimensions
Data Scraping	Data from new accounts, CRM portals, etc.	No. of account, No. of disputes	Date, Quarter, Month
Data Sorting	Data accuracy, consistency	No. accounts with insufficient data	Quarterly
Reports	Legit report	System overview, Product issues, Product sales	Quarter, Month

5. Dimensional Model:



6. Data Quality:

a. Error checking:

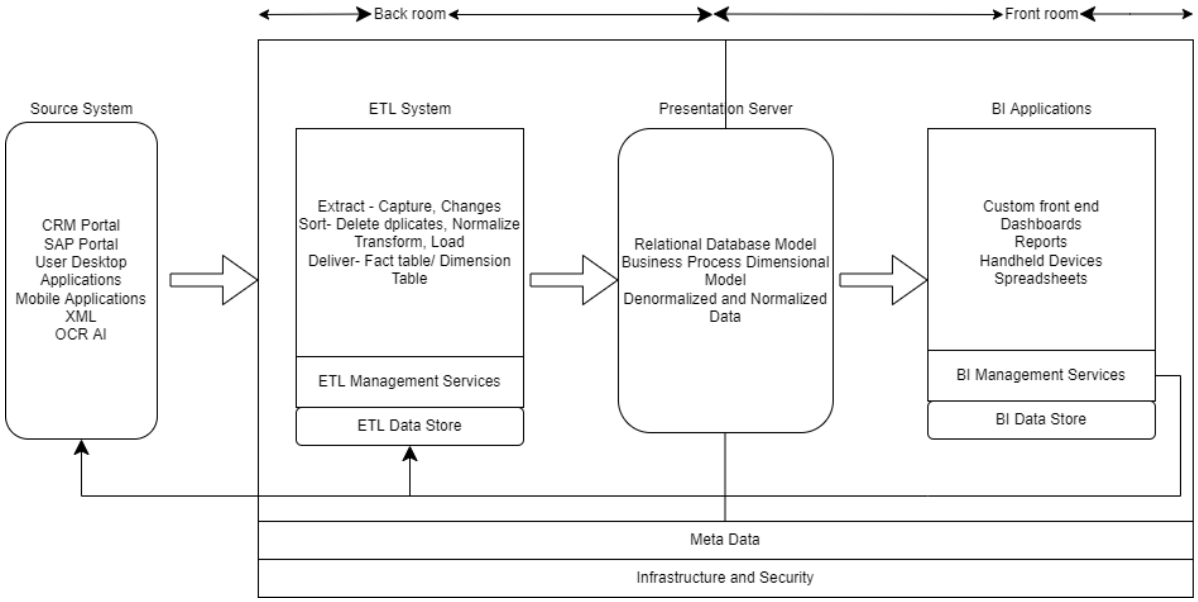
Table Name	Column Name	Alias/Surrogate	Data Type	Length	Primary Key	Foreign Key	Index Used	Constraints	Example
New_Account_Dim	Account_No		Int	16	Yes		NA	Not Null	3.26599E+17
	First_Name		varchar	50			NA	Not Null	Jose
	Last_Name		varchar	50			NA	Not Null	Victor
	Customer_Address		varchar	100			NA	Not Null	123, N Waveland
	City		varchar	50			NA	Not Null	Chicago
	State		varchar	50			NA	Not Null	Illinoi
	Zip		Int	6			NA	Not Null	60645
	Phone		Int	10			NA	Not Null	1236547896
	Documents_Submitted		varchar	50			NA	Not Null	Lease Doc
	Customer_SSN		Int	10			NA	Not Null	3334446598
	Services		varchar	50			NA	Not Null	Online Banking
Disputes_Dim	Dispute_ID		Int	10	Yes		NA	Not Null	1236547896
	Account_No		Int	16			NA	Not Null	3.26599E+17
	Customer_ID		Int	16			NA	Not Null	3.26599E+17
	Dispute_Time		varchar	10			NA	Not Null	11:00 PM
	Complaint_Time		varchar	10			NA	Null	11:00 PM
	Dispute_Type		varchar	10			NA	Not Null	Transaction
	Dispute_Location		varchar	50			NA	Not Null	Chicago, downtown
	Dispute_End		varchar	10			NA	Not Null	02-02-2022
Insurance_Dim	Insurance_Police_No		Int	10	Yes		NA	Not Null	1236547896
	Policy_Holder_Name		Varchar	50			NA	Not Null	Jose
	Policy_Holder_Address		Varchar	100			NA	Not Null	123, N Waveland
	Insurance_Start_date		Date				NA	Not Null	02-02-2022
	Insurance_End_date		Date				NA	Not Null	02-02-2022
	Insurance_Type		Varchar	50			NA	Not Null	Car Insurance
Loans_Dim	Loan_Policy_No		Int	10	Yes		NA	Not Null	1236547896
	Loan_Account_No		Int	16			NA	Not Null	3.26599E+17
	Borrower_FName		varchar	50			NA	Not Null	Jose
	Borrower_LName		barchar	50			NA	Not Null	Victor
	Borrower_Address		varchar	100			NA	Not Null	123, N Waveland
	Issuing_Bank_ID		Int	10			NA	Not Null	1236547896
	Loan_Type		varchar	50			NA	Not Null	Home
Bank_Dim	Bank_ID		Int	10	Yes		NA	Not Null	1236547896
	Branch_Code		Int	10			NA	Not Null	1236547896
	Bank_City		Varchar	50			NA	Not Null	Chicago, downtown
	Bank_Zip		Int	10			NA	Not Null	60645
	Bank_Phone		Int	10			NA	Not Null	1236547896
	Accounts_Count		Int	10			NA	Not Null	1000
	ATM_Count		Int	10			NA	Not Null	500
	Staff_Count		Int	10			NA	Not Null	100000
	Revenue		Int	100			NA	Not Null	1000000000
Calander_Dim	Date_ID		Int	10	Yes		NA	Not Null	12
	Date		Int	2			NA	Not Null	12
	Month		varchar	10			NA	Not Null	March
	Year		Int	4			NA	Not Null	1994
	Quarter		varchar	10			NA	Not Null	4th
ATM_Dim	ATM_ID		Int	10	Yes		NA	Not Null	1236547896
	Branch_ID		Int	10			NA	Not Null	1236547896
	Branch_Name		varchar	50			NA	Not Null	Downtown
	Location		varchar	100			NA	Not Null	123, N Waveland
	Transactions		Int	100			NA	Not Null	120
	ATM_Bal		Int	100			NA	Not Null	100000
	ATM_Issues		boolean				NA	Null	yes
Fact_report	Report_ID		Int	10	Yes		NA	Not Null	1236547896
	Date_ID		Int	10		Yes	NA	Not Null	1236547896
	Bank_ID		Int	10		Yes	NA	Not Null	1236547896
	Loan_Policy_No		Int	10		Yes	NA	Not Null	1236547896
	ATM_ID		Int	10		Yes	NA	Not Null	1236547896
	Account_No		Int	16		Yes	NA	Not Null	3.26599E+17
Fact_Status	Status_ID		Int	10	Yes		NA	Not Null	1236547896
	Account_No		Int	16		yes	NA	Not Null	3.26599E+17
	Dispute_ID		Int	10		yes	NA	Not Null	1236547896
	Insurance_Policy_No		Int	10		yes	NA	Not Null	1236547896
	Loan_Policy_No		Int	10		yes	NA	Not Null	1236547896
Agg_Fact_Report	Agg_Fact_Report_ID		Int	10	Yes		NA	Not Null	1236547896
	Date_ID		Int	10		Yes	NA	Not Null	1236547896
	Report_ID		Int	10		Yes	NA	Not Null	1236547896
	Status_ID		Int	10		Yes	NA	Not Null	1236547896

b. Standards:

Standardization Table Area of Applicability	Standard	Example
Date	mm-dd-yyyy	02-02-2022
Account_No	1234567894	123466987
Name	Name	Jose
SSN	123-123-1234	656-987-6545

7. Architectural Deliverables:

a. End-to-End Architecture:



b. Source System Table:

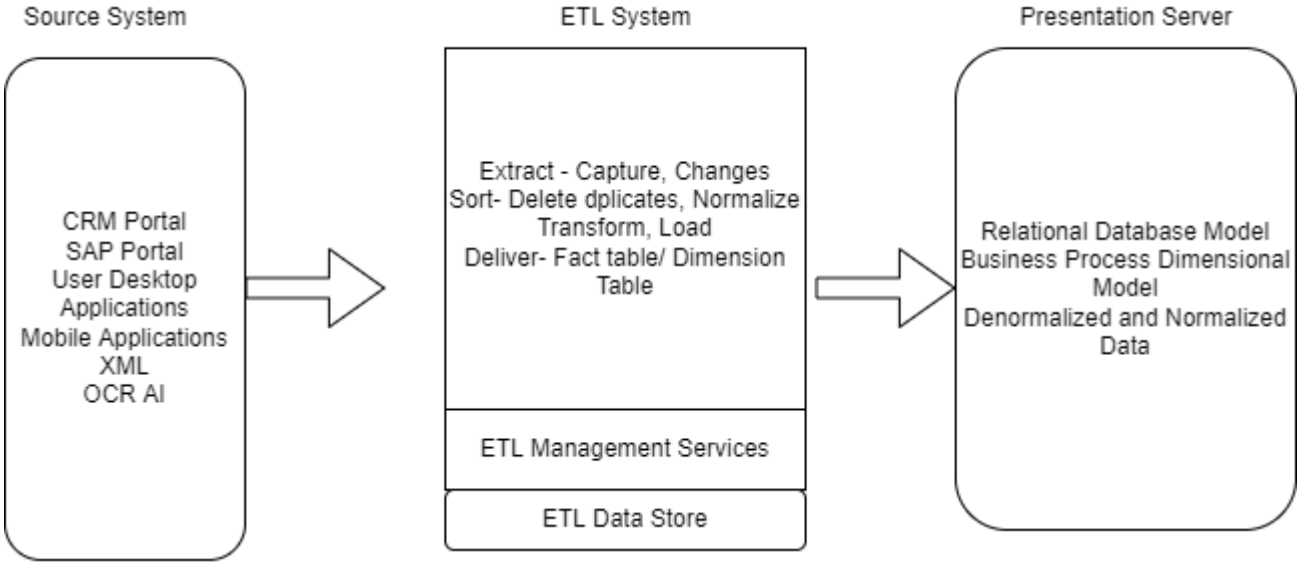
Source	Business Owner	IS Owner	Platform	Location	Description
CRM	William Lee	Fred Kicks	User Application	HQ	POS data from new user applications
Online Portal	Faith Mining	Jose Richards	Website	HQ Mumbai	Online user data from websites
SAP	Isabella Terring	Tylor Frost	SAP Application	Pune	Data from SAP application portal
Complaints	Zara	Sampath Reddy	User Application	HQ	Data from complaints portal
ATM	Deepanjan Rane	Anuj Malve	User Application	HQ	ATM records for different branches
Customer Service	John Roberts	Raj Naik	Windows	HQ	Customer interaction data for various service request

c. ETL Components:

Extract- This is in context with gathering or collecting data from the source which can be different applications or platforms such as CRM portal etc. Sources can also consist of relational databases, XML files, etc. this data is captured and further processed for usability in some cases data can be changed to meet the requirements or else deleted which not meeting the constraints. Or can be sent for reconsideration.

Sort- In this, data is sorted using the mechanism to delete duplicate data. It can also be stated as a cleansing of data. Removing duplicates and working on normalization.

Transformation and Loading- This is reshaping the data to meet the destination requirements and loading data into the destination data format or requirements. It is also one of the reshaping processes like sorting. Loading of data can be done in a variety of formats where the user needs the data for further implementation, usually, the data is stored in database or data stores.



d. BI Tools:

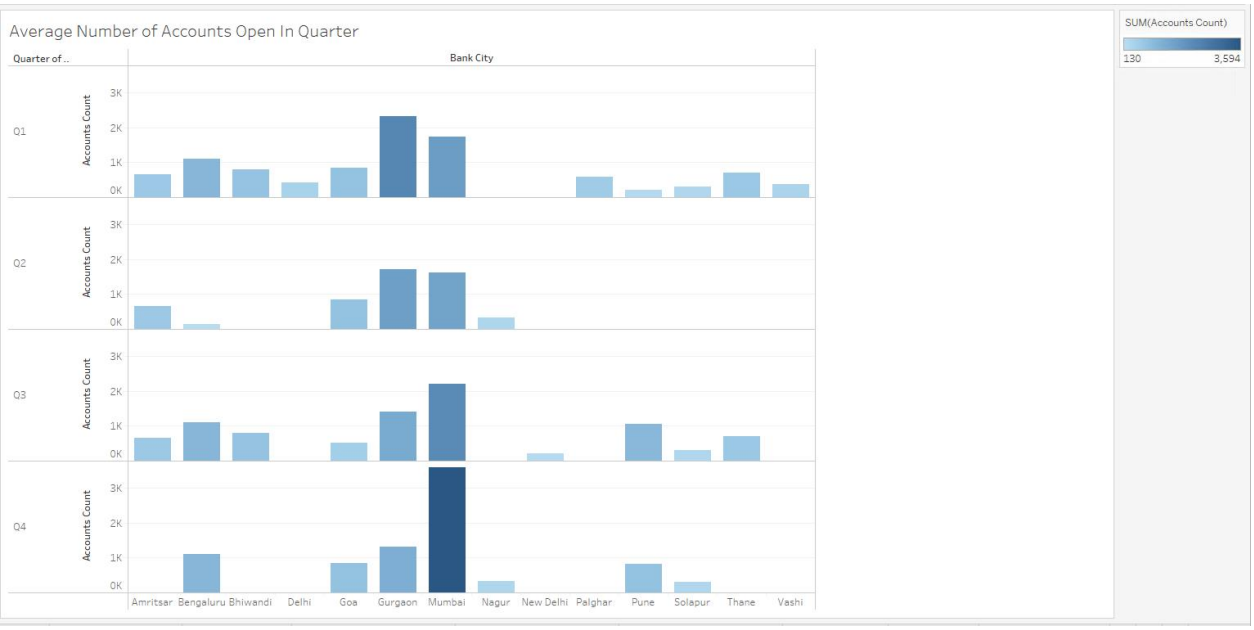
BI Tool	Description
Tableau	Simple and easy way to visualize data. One of the commonly used BI tools. It is a paid tool but its public version is free to access with limited features.
Power BI	Another very well-known BI tool in the market. It has more functionality than Tableau. It is a paid and licensed product.

e. BI User table:

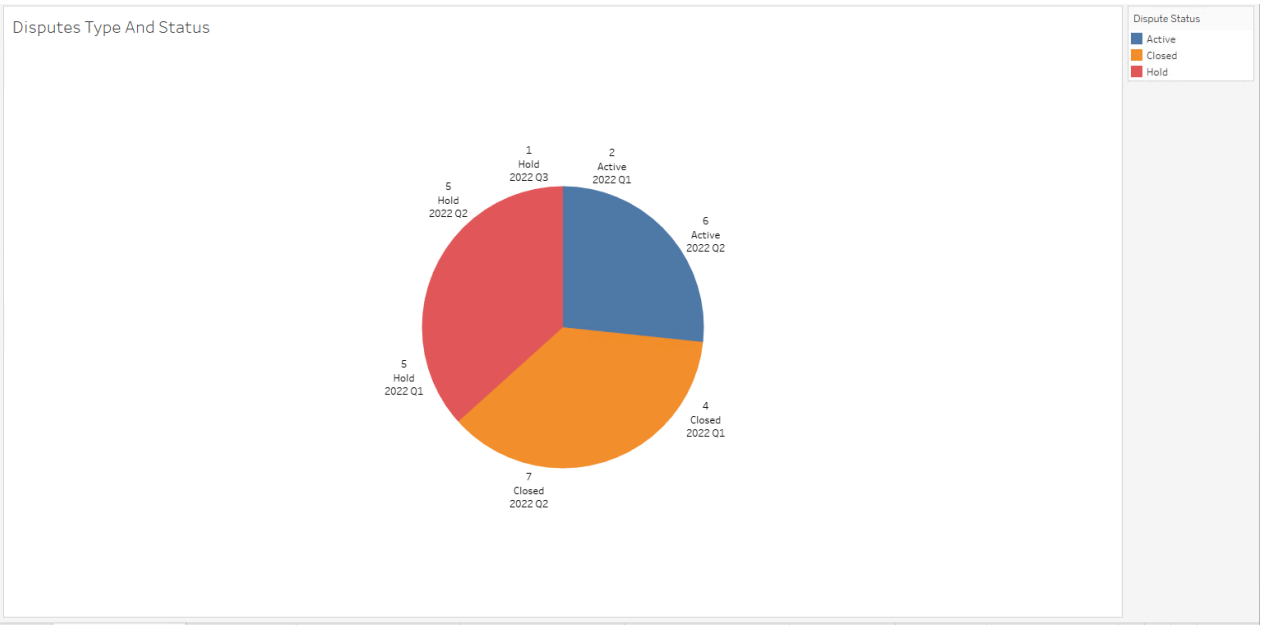
User	How
Product Manager	One can check their product and service-related data on the reports dashboard on Tableau.
Branch Managers	One can check different status reports on the status dashboard in Tableau.
CEO and other higher authorities	Can view consolidated reports on the aggregate reports dashboard.

8. Dashboards and Reports:

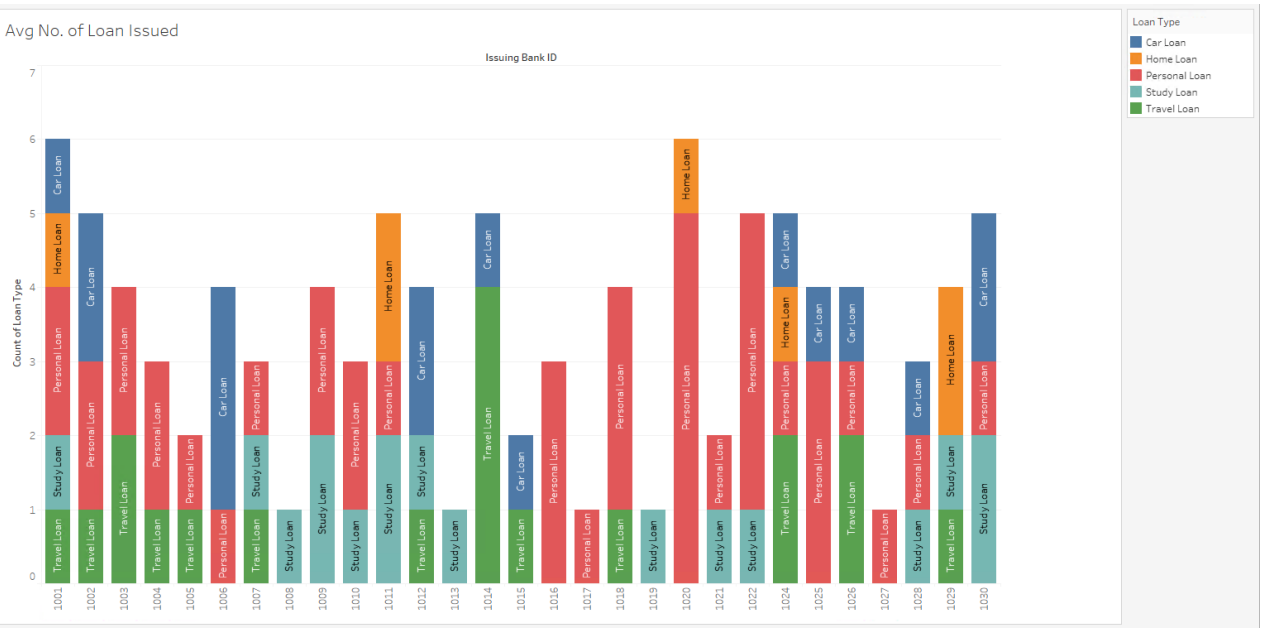
Reports:



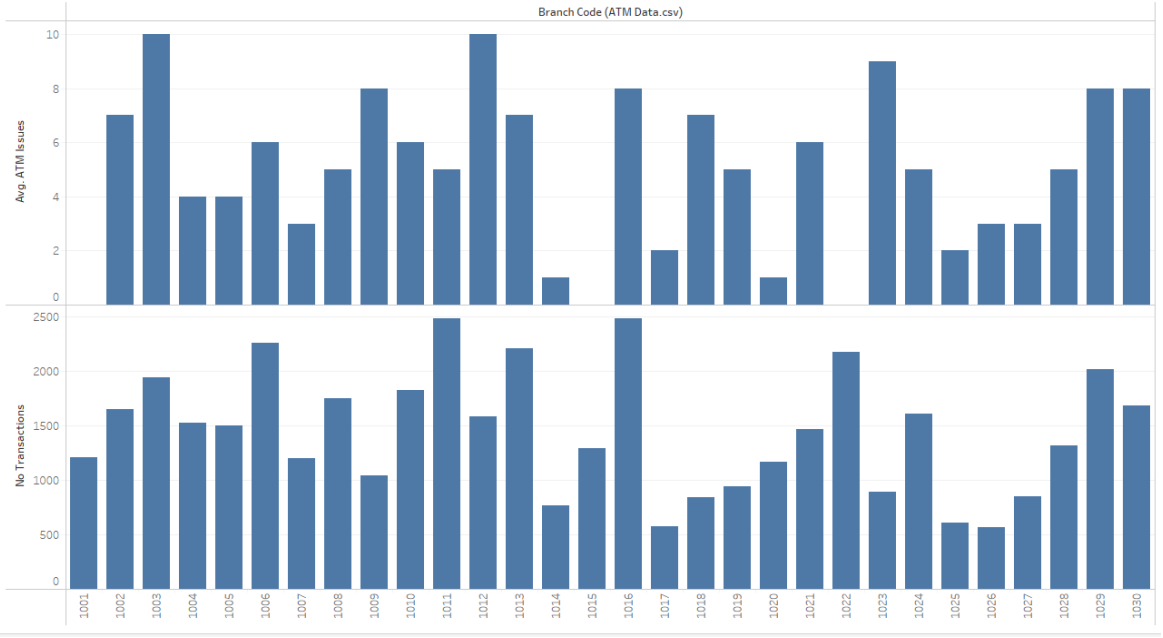
Disputes Type And Status



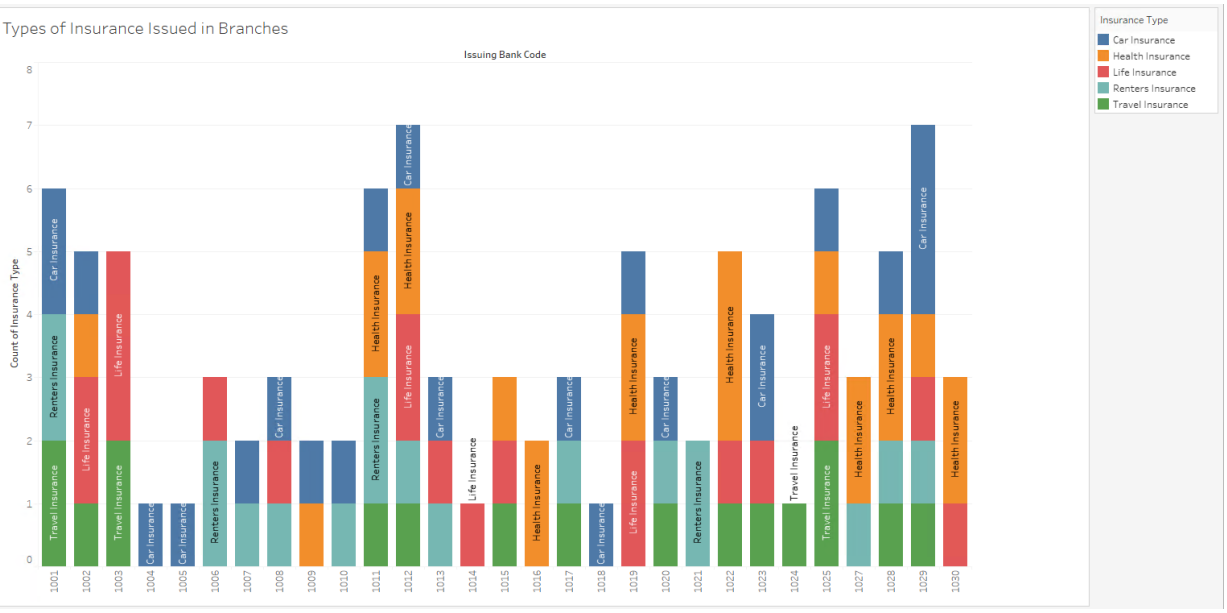
Avg No. of Loan Issued



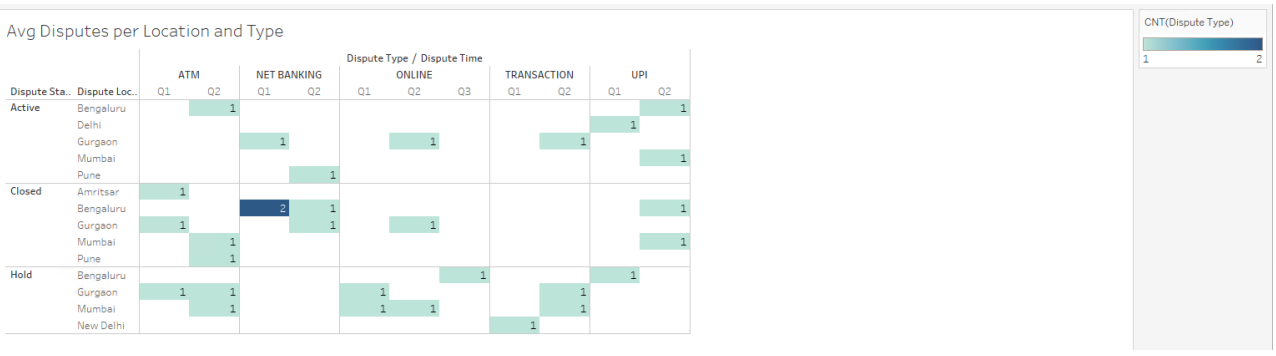
Avg No of ATM Transactions and Issues

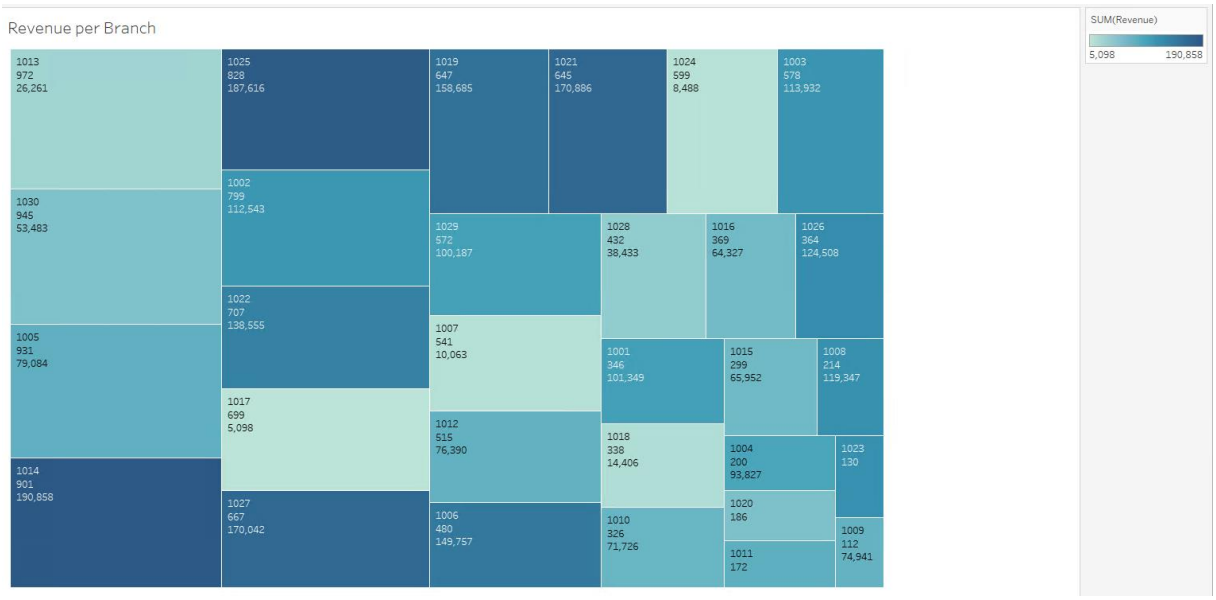


Types of Insurance Issued in Branches

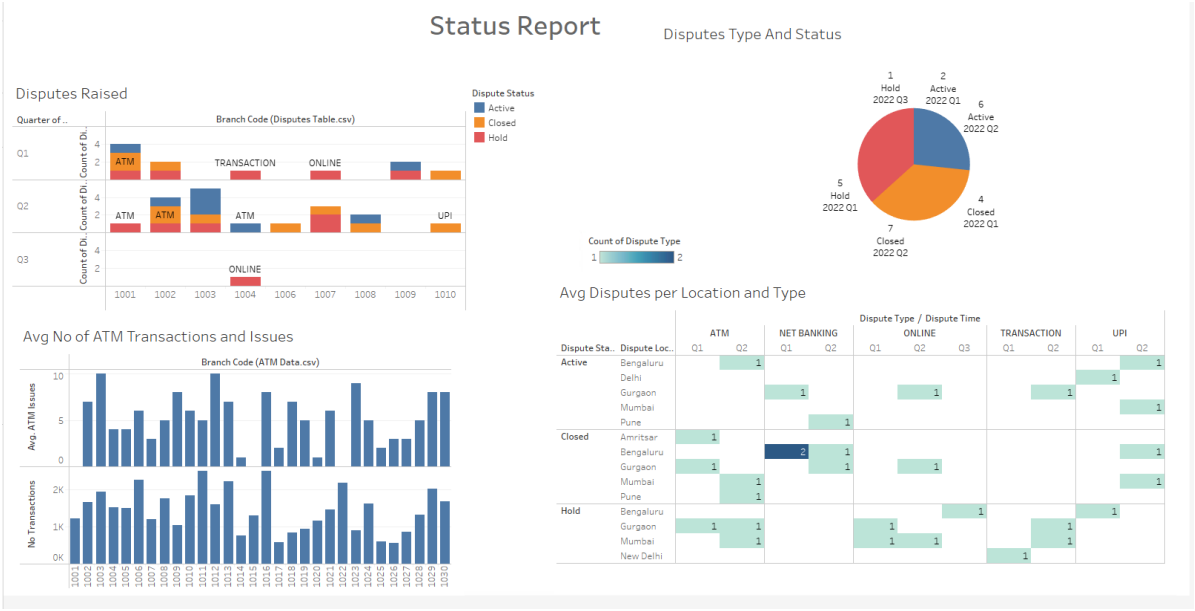


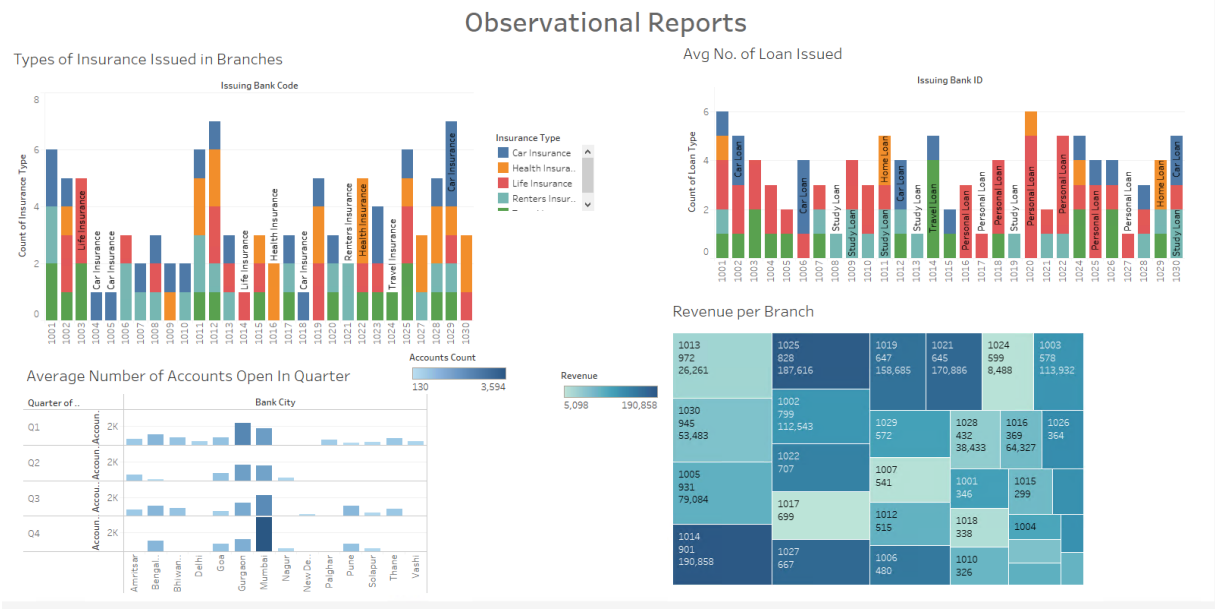
Avg Disputes per Location and Type



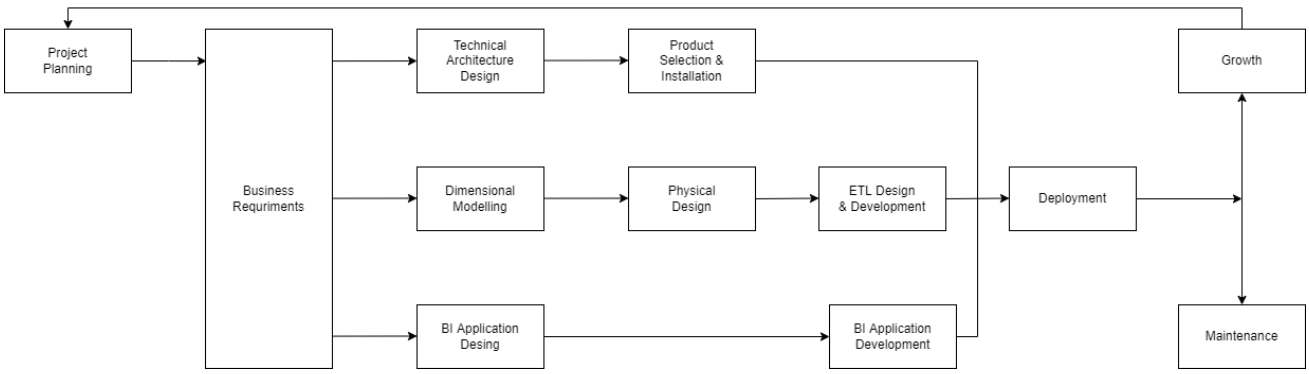

Revenue per Branch
SUM(Revenue)
5,098
190,858
1013 972 26,261
1025 828 187,616
1019 647 158,685
1021 645 170,886
1024 599 8,488
1003 578 113,932
1030 945 53,483
1002 799 112,543
1029 572 100,187
1028 432 38,433
1016 369 64,327
1026 364 124,508
1005 931 79,084
1022 707 138,555
1007 541 10,063
1001 346 101,349
1015 299 65,952
1008 214 119,347
1017 699 5,098
1012 515 76,390
1018 338 14,406
1004 200 93,827
1023 130
1014 901 190,858
1027 667 170,042
1006 480 149,757
1010 326 71,726
1020 186
1009 112 74,941
1011 172

Dashboards:


Status Report
Disputes Raised
Dispute Status
Active
Closed
Hold
Quarter of ..
Branch Code (Disputes Table.csv)
Count of Di..
Q1
Q2
Q3
ATM
TRANSACTION
ONLINE
UPI
1001 1002 1003 1004 1006 1007 1008 1009 1010
Disputes Type And Status
1 Hold 2022 Q3
2 Active 2022 Q1
6 Active 2022 Q2
5 Hold 2022 Q1
4 Closed 2022 Q1
7 Closed 2022 Q2
Count of Dispute Type
1
2
Avg No of ATM Transactions and Issues
Branch Code (ATM Data.csv)
Avg ATM Issues
No Transactions
Avg Disputes per Location and Type
Dispute Type / Dispute Time
ATM
NET BANKING
ONLINE
TRANSACTION
UPI
Q1
Q2
Q3
Dispute Sta..
Dispute Loc..
Active
Closed
Hold
Bengaluru
Delhi
Gurgaon
Mumbai
Pune
Amritsar
New Delhi


Observational Reports
Types of Insurance Issued in Branches
Issuing Bank Code
Count of Insurance Type
Insurance Type
Car Insurance
Health Insura..
Life Insurance
Renters Insur..
Average Number of Accounts Open In Quarter
Accounts Count
130
3,594
Quarter of ..
Bank City
Q1
Q2
Q3
Q4
Amritsar
Bangal..
Bhiwan..
Delhi
Goa
Gurgaon
Mumbai
Nagur
Navi De..
Palghar
Pune
Solapur
Thane
Vashi
Avg No. of Loan Issued
Issuing Bank ID
Count of Loan Type
Car Loan
Home Loan
Study Loan
Travel Loan
Personal Loan
Revenue per Branch
Revenue
5,098
190,858
1013 972 26,261
1025 828 187,616
1019 647 158,685
1021 645 170,886
1024 599 8,488
1003 578 113,932
1030 945 53,483
1002 799 112,543
1029 572
1028 432 38,433
1016 369 64,327
1026 364
1005 931 79,084
1022 707
1007 541
1001 346
1015 299
1017 699
1012 515
1018 338
1004
1014 901 190,858
1027 667
1006 480
1010 326

9. Kimball's Roadmap:



Project Planning:

ABC Bank is implementing a data warehouse to efficiently gather data from various sources and analyze and evaluate the same, which will enable them for future growth. This will also help them in knowing their key focus in the market and what more products they can introduce in the market.

Business Requirements:

This will collaboratively work with different departments of the bank to closely check the business outputs from each department and collect them as resources for the project. As this project is working on overall organizational development, every single piece of data needs to be collected and understood for further implementation.

Dimensional Modelling:

It is a detailed overview of the entire business broken into small parts so that it is better understood by everyone working on the project. Dimensional modeling enables us to understand the data flow and value throughout the organization. This also enables us to define all tables and databases we are required to design for the implementation. Majorly this defines our KPI's facts and dimensions.

Technical Architecture Design:

This architecture defines the project or process flow once it is implemented. It begins will collecting data from the source system, then the data is sent through ETL and is further uploaded into the database model, which is then used by various BI tools for further analysis.

BI Application:

It is a presentation layer wherein we will use different BI tools for analyzing our data. The data from various source systems go through ETL and then it's stored in databases from where the BI tools get their input and this data is analyzed and worked up for further decisions.

Deployment:

This is the final pathway where the reports or the analysis are released to the management team or authorities wherein they make this final decision about the business and gain fruitful information from the raw data that was previously uploaded to the system.

Growth and Maintenance:

The process is constantly monitored for changes and upgrades as the market demands. It is maintained every now and then due because this implementation performs necessary decisions for the organization. Producing or introducing best-in-market products for its customer is step towards growth.