DataVerse

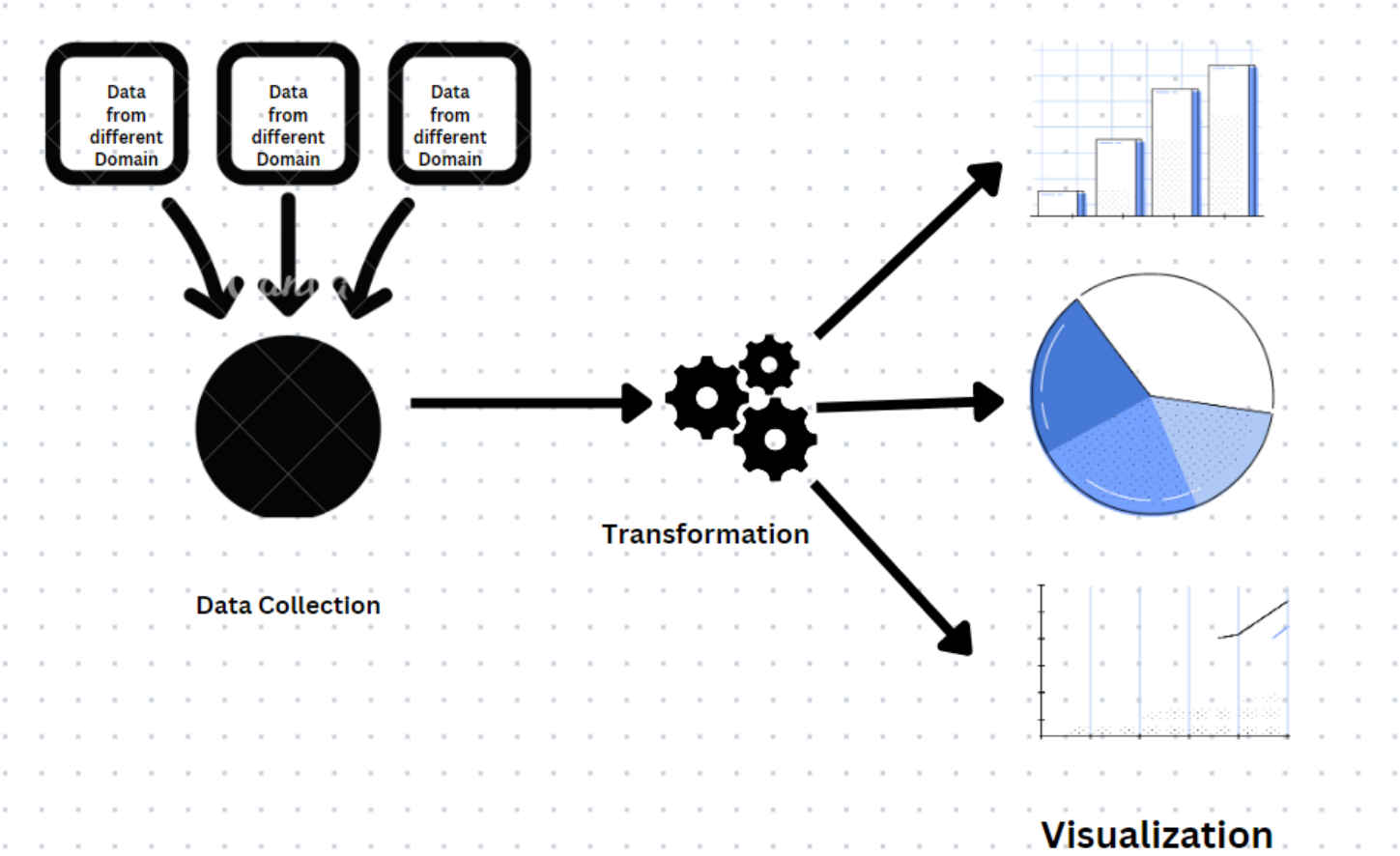
Data visualization for Finance

Group 6

Aadarsha Chapagain (C0825975)
Rajasekhar katta (C0833766)
Venkata sai manikanta Ponakala (c0833772)
Rishi Phaneendra Varma Bhupathi Raju(C0825285)

Contents

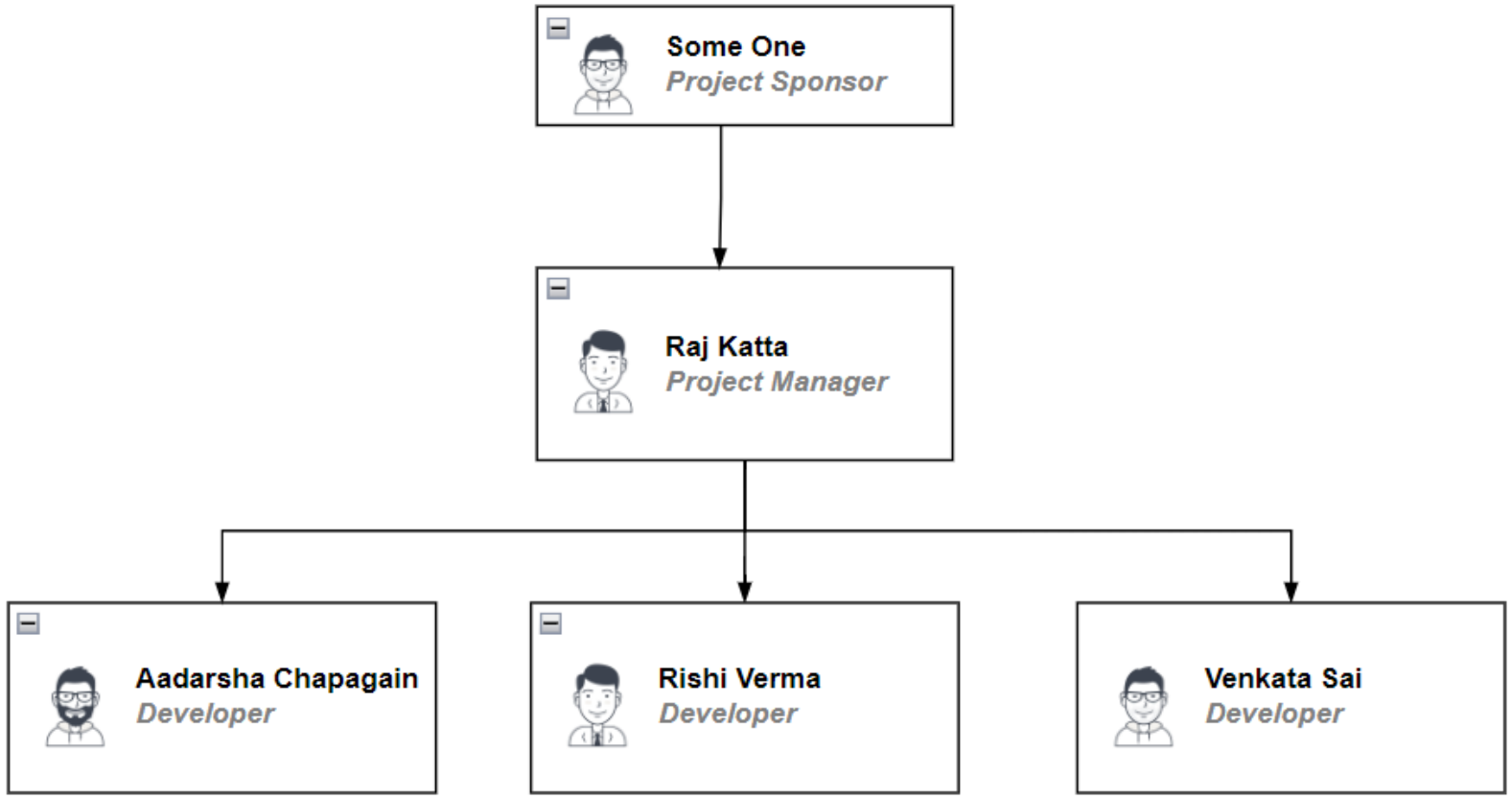
What is DataVerse



Project Title and description

- Financial institutions are struggling to get insights from their data
- Excel is not enough to handle huge amount of data
- Just Accumulating data won’t add any value to organization
- Companies utilizing data as an information assets are ahead of the game
- **Dataverse** is a data visualization project where data from financial institutions will be collected, transformed and visualized

Project Team



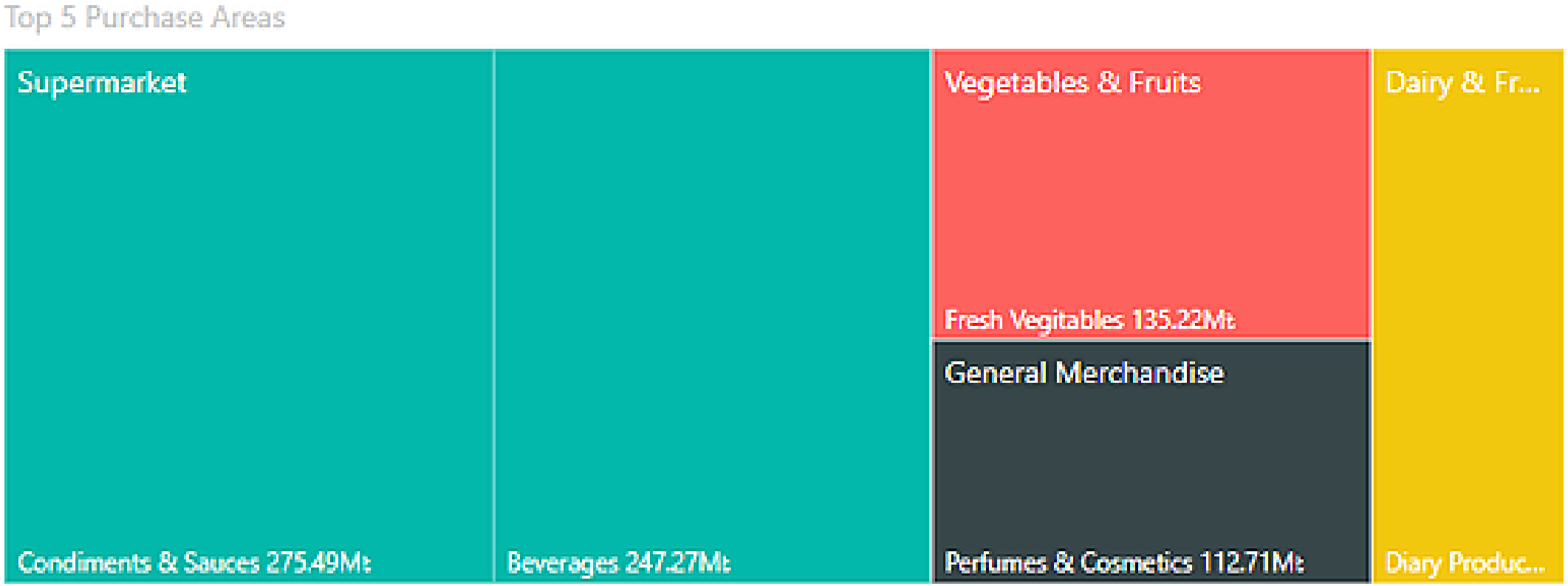
Business Case

- This project targets banking customers when and how they are likely to be most amenable to making a purchase.

- This provides information to help optimize banking marketing efforts. This knowledge can involve anything. Client may design a new product or service and know when and where to market them.

- Additionally, if a client experiencing a slowdown in sales, the data will provide insights into what is wrong.

Credit card spending Visualisation



Technology & Resource

- This project leverages Microsoft Azure Public Cloud services as Pay-as-you-Go subscription to avoid capital investments and operational maintenance.

- Rajasekhar katta will be the Project Manager and Data modeler.
- Rishi Varma will be appointed as Data Engineer.
- Aadarsha Chapagain will be an Azure Devops developer.
- Venkata Sai Manikanta will be PowerBI developer.

WBS

	STORE	
1	Source Systems, data types, formats, schemas	List out all the source systems that will be pumping data to the storage. For each of the source system define, frequency of data, data type, schema, file type, data model all data attributes
2	Target systems that extract/access data	List out all the target systems will extract/access data from storage. For each of the target systems define frequency of data, data type, schema, file types, data modeol all data attributes
3	Data growth, storage costs forecast & estimation	Define data growth for short term < 6 months, 1 year and 3 years, regardless of cloud/data center storage. Identify storage costs upfront based on cloud option selected, indetify areas of improvement
4	Data Versioning	Define the versioning mechanism for the incoming data and data that will be stored in storage long term for access. Define what does a version mean and when a verion changes, what backward compatiblity is suppoted
5	Data Layout	Define the layout structure of the data for ingested/stored/processed/accessed data
6	Metadata management	Identify tool for cataloging and metadata management. Verify and validate if source and target systems are ineroperable with the catalog and metadata tool. Define standard policies for tagging data objects, ignoring this at the beginning will create confusion with data from multiple source, its important to define this upfront what does a tag look like, will it be source name, or target name or process that is touched the data etc... and all the processes automatic and manual need to be consistent with this definition Define backup and restore strategies for catalog/metadata
7	Encryption at rest	Define encryption strategy for the data at rest, will it be client-side /server-side encryption and if server-side encryption and how to do key management
8	Data Archiving and Retrieval	Define Storage policies for data archiving, the time period, duration all the attributes on which data would be archived to cheaper tiered storage if on cloud. Define retrieval polcies this is important particularl if on cloud, the costs can spiral based on retrieval access patterns, so its also important to define how and when a data retrieval can be made.
9	Data Retention	Define how a long a particula data set from a source and used by a target is retained or not deleted. As much as data archival to tiered storage helps in reducing costs and data can be kept perpetually, its also important not have junk that is pushed to some cheap storage that would be never be required to use. Define compliance policies for data retention
10	Data Access patters	How do you plan to access data API -> for automation, JDBC Access -> SQL Access, Data Export -> Bulk Data Tools to push data to external systems, what kind of tools, how the tools align with security mechanisms
11	Storage Optimization	Optimize storage for cost, forecast data growth and monitor storage metrics, use cloud provided services to get optimization insights and Automate report generation of storage metrics
12	Authentication & Authorization	Define authentication & Authorization processes and mechanisms for Automated Services, Users, Tools Automate report genereration of storage/data access for auditing
13	Disaster recovery and geo redundancy	Define Disaster recovery requirements Define policies for data replication including geo redundancy replication Define tools and process for disaster recovery Define testing/simularion strategy for disaster recovery testing Automate disaster recovery testing

Key Stakeholder List

- Customers
- Project Manager
- Project Team Members
- Company Executives
- Project Sponsor
- Steering Committee
  - An Advisory group providing guidance on key decision's on sponsor, executives and key stakeholders

Stakeholder Requirements as known

- Business unit: Enterprise Business Intelligence (EBI)
- Business Operations
- Customers
- Subject Matter Experts

Product Description

- This project aims to help the clients by reaching the banking customers and shows when they are most likely to be open to making a purchase.

- By creating the PowerBI dashboard with all the required data the client can access lot of information about their customers spending patterns in a simple way.

- This will provide the information about the customers purchase history, their spending limits and many more.

Key Deliverables

- Customers spending patterns
- Big purchases in last 6 months
- How many times they exceed the credit limit in past 2 years
- Repaying on time or not
- Their top 5 spending product categories
- All these insights will be presented on PowerBI Dashboard

References

- https://learn.microsoft.com/en-us/power-bi/fundamentals/service-get-started
- https://learn.microsoft.com/en-us/power-bi/create-reports/service-dashboard-create
- https://truelayer.com/blog/data-chain-retrieving/
- https://docsumo.com/blog/best-bank-data-extraction-software
- https://www.inetco.com/use-cases/credit-and-debit-card-analytics/