A Major Project Synopsis on

DATA ANALYSIS PROJECT

Submitted to Manipal University, Jaipur

Towards the partial fulfillment for the Award of the Degree of

MASTER OF COMPUTER APPLICATIONS

2023-2025

by

DHARMESH SINGH

23FS20MCA00115

Under the guidance of

Dr. Chandrashekhar Patel

Department of Computer Applications

School of AIML, IoT&IS, CCE, DS and Computer Applications

Faculty of Science, Technology and Architecture

Manipal University Jaipur

Jaipur, Rajasthan

2025

Table of Contents

1. Introduction

2. Motivation

3. Statement of Problem

4. Methodology/ Planning of work

5. Data flow diagram

6. Requirements for proposed work

7. Future Scope

8. Bibliography/References

Introduction:

The banking sector is dealing with an unprecedented flood of structured and unstructured data in the current era of digital finance. This data comes from a variety of sources, such as digital interactions, loan processing systems, credit evaluations, customer transactions, and investment tracking platforms. For banks hoping to prosper in a changing market, using this data to extract actionable insights is not only a competitive advantage but also a fundamental necessity.

The goal of this project, "Holistic Banking Dashboard Utilizing Power BI and Python," is to create, develop, and implement an analytics-driven, interactive dashboard solution that combines the powerful statistical analysis and data manipulation features of Python with the real-time visualization capabilities of Microsoft Power BI. This program focuses on turning unstructured financial data into informative KPIs and reports so that all organizational levels may make data-driven decisions.

Comprehensive data on customer profiles, loan disbursements, deposit summaries, engagement levels, income segmentation, and bank performance metrics are all included in the scope. Strategic evaluations including risk profiling, client segmentation, and resource allocation optimization will be supported by the solution.

.

1. Motivation

The main reason for starting this research is to close the analytical gaps that exist in conventional banking systems. Financial organizations frequently use antiquated, inflexible IT platforms that offer disjointed perspectives of how businesses operate. In general, these systems lack the flexibility to adjust to contemporary banking issues including real-time monitoring, predictive risk modelling, and analytics on client involvement.

The main motivators are:

Data Centralization and Visualization: To present a comprehensive picture of banking operations, combine several datasets into a single platform.

Risk Mitigation: Profile consumer behavior across multiple demographics using data analytics to predict credit risks and default trends.

Strategic Intelligence: Provide bank executives with dashboards that provide useful information about client activity levels, deposit trends, and financial KPIs.

Efficiency and Automation: To free up data experts to concentrate on innovation, automate routine reporting cycles and analytical processes.

Improved Customer Understanding: Use analytics to analyse consumer engagement and behavior in order to provide tailored financial services.

The project is in line with worldwide trends toward the transition of digital banking, where business intelligence and artificial intelligence capabilities are becoming essential to daily operations.

2. Problem Statement:

The prime motivation behind undertaking this project lies in bridging the analytical gaps prevalent in traditional banking systems. Financial institutions often rely on dated, rigid software platforms that provide fragmented views of business operations. These systems are generally ill-equipped to adapt to modern banking challenges such as predictive risk modeling, real-time monitoring, and client engagement analytics.
Key motivational drivers include:

Data Centralization and Visualization: Integrate disparate datasets into a unified environment to provide a holistic view of banking operations.

Risk Mitigation: Use data analytics to anticipate default patterns and credit risks by profiling customer behavior across various demographics.

Strategic Intelligence: Equip bank executives with dashboards that offer actionable insights into financial KPIs, deposit trends, and client activity levels.

Efficiency and Automation: Automate repetitive analytical tasks and reporting cycles to allow data professionals to focus on innovation.

Enhanced Customer Understanding: Dissect customer behavior and engagement through analytics to offer targeted banking services.

The initiative aligns with global trends toward digital banking transformation, where artificial intelligence and business intelligence tools are becoming integral to operations.

3. **Methodology/ Planning of work:**

    The project methodology follows an iterative, modular design process incorporating both exploratory and confirmatory data analysis techniques:

    Phase 1: Data Collection and Source Integration
    - Import datasets such as Banking.csv, clients.csv, and Banking.xlsx containing over 50 fields related to customer profiles, financial metrics, and transactional records.
    - Utilize Power BI's built-in connectors and Python's pandas to ingest and validate data.

    Phase 2: Data Exploration and Analysis (Python)
    - Use Python libraries (pandas, seaborn, matplotlib) for cleaning, exploration, and visualization.
    - Identify outliers, missing data, and correlations between features.
    - Generate descriptive statistics and distributions for fields such as income, deposit volume, and loan types.

    Phase 3: Data Transformation and Feature Engineering
    - Add calculated fields:
        - Engagement Days: Difference between account opening date and current date.
        - Income Band: Classification into Low, Mid, and High-income segments.
        - Processing Fees: Derived from fee structure using nested logic statements.
    - Normalize data formats and encode categorical variables as required.

    Phase 4: Data Modelling and Schema Creation in Power BI
    - Establish entity relationships using primary and foreign keys (Client ID, Bank ID).
    - Create a star schema structure with dimensional tables (Client, Bank, Advisor, Time) and fact tables (Transactions, Loans).

    Phase 5: DAX Measures and KPI Formulation
    - Develop calculated fields and KPIs using DAX:
        - Total Loan, Total Deposit, Total Fees, Bank Engagement, Client Retention Index

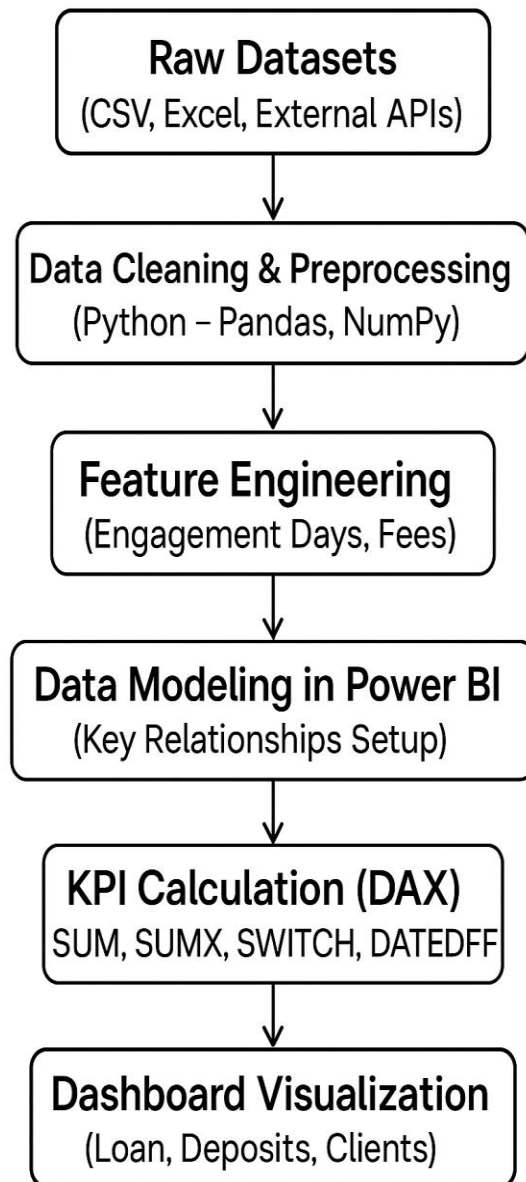- o Use SUMX, SWITCH, CALCULATE, and DATEDIFF for advanced logic.

Phase 6: Dashboard Design and Visualization
- Construct dashboards for:
  - o Loan Distributions by demographic filters (gender, income band, nationality).
  - o Deposit Patterns across account types (checking, savings, foreign currency).
  - o Engagement Trends: Timeline plots for client activity and growth metrics.
  - o Financial Summaries: Aggregated visualizations of fees, deposits, loans.

Phase 7: Reporting and Documentation
- Generate a presentation using PowerPoint and a final documentation report summarizing all insights, methodologies, and conclusions.

4. Data flow diagram /DFD:

```
┌─────────────────────────────┐
│       Raw Datasets          │
│  (CSV, Excel, External APIs)│
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│ Data Cleaning & Preprocessing│
│   (Python – Pandas, NumPy)  │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│     Feature Engineering     │
│   (Engagement Days, Fees)   │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│  Data Modeling in Power BI  │
│  (Key Relationships Setup)  │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│    KPI Calculation (DAX)    │
│  SUM, SUMX, SWITCH, DATEDFF │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│    Dashboard Visualization  │
│   (Loan, Deposits, Clients) │
└─────────────────────────────┘
```

5. Requirements for proposed work:

To successfully implement this project, the following hardware and software resources are required:

Software Tools:
- Power BI Desktop (latest stable release)
- Jupyter Notebook (via Anaconda Distribution)
- Microsoft Excel or Google Sheets
- Python 3.10 or later
-

Python Libraries:
- pandas: Data manipulation and preprocessing
- matplotlib and seaborn: Data visualization and trend analysis
- scikit-learn: Future integration for ML-based prediction
-

Hardware Configuration:
- Processor: Intel i7 or AMD Ryzen 7
- Memory: Minimum 16 GB RAM
- Storage: 512 GB SSD
- GPU: Optional (useful for future deep learning integration)

6. Future Scope:

This foundational project opens the door to a wide spectrum of future enhancements that can increase the breadth, depth, and usability of the dashboard system:

A. Predictive Analytics Integration
- Build predictive models using logistic regression, decision trees, or neural networks.
- Automate risk assessment scores for new loan applications.

B. Real-time Dashboarding
- Integrate Power BI with streaming data from APIs or banking databases.
- Allow real-time alerts and live KPI monitoring.

C. Client Segmentation and Personalization
- Use clustering algorithms to categorize clients into behavioural groups.
- Generate dynamic client personas for targeted product recommendations.

D. Regulatory Compliance and Audit Trail
- Design audit dashboards to ensure adherence to regulations like AML, KYC.
- Implement data lineage and traceability tools.

E. Cross-institution Benchmarking
- Compare metrics across different financial institutions.
- Create dashboards that rank bank branches by performance.

F. Mobile and Web App Deployment
- Host dashboards in the Power BI service or embed them in custom portals.
- Optimize dashboards for use on tablets and smartphones.

7. Bibliography/References:

1. Microsoft Power BI Official Documentation –
   https://docs.microsoft.com/en-us/power-bi/

2. Python Data Analysis Libraries:

   - pandas: https://pandas.pydata.org/

   - seaborn: https://seaborn.pydata.org/

   - matplotlib: https://matplotlib.org/

3. DAX Reference and Syntax – https://learn.microsoft.com/en-us/dax/

4. Reserve Bank of India (RBI) Regulatory Framework
5. Financial Risk Management Research – Journal of Banking Analytics, 2022.