**Technical Documentation Banking**

**Title of the project:**

* Comprehensive Financial Behavior Analysis Using User Demographics

**Abstract:**

* This project analyzes user demographics and financial transactions to provide actionable insights. By leveraging various data preprocessing, EDA, and feature engineering techniques, the project uncovers patterns in user behavior, income distribution, cash-out frequency, and support interactions. The findings are visualized through Power BI dashboards, aiding stakeholders in making informed decisions.

**Table of Contents:**

* Introduction
* Project Overview
* Data Collection
* Data Analysis
* Implementation
* Evaluation
* Conclusion
* References
* Appendices

**1. Introduction:**

* To analyze user demographics and financial transactions for actionable insights. It helps stakeholders understand user behavior and economic patterns to enhance decision-making.

 Goals **and Objectives:**

* Analyse user demographics (gender, region, income).
* Investigate financial behaviours (cash-outs, tips).
* Evaluate user support interactions.
* Visualize findings through Power BI dashboards.

**2. Project Overview:**

* **Scope:** Analysis of user demographics, financial transactions, and support interactions.
* **Data Sources:** Synthetic datasets for users and transactions, including user demographics, transaction details, and support interactions.
* **Context:** The project was undertaken to provide comprehensive insights into user and financial behavior for a financial services company.
* **Constraints and Limitations:**
  + Lack of real-world data may limit the applicability of findings.
  + Synthetic data may not perfectly represent real user behavior.

**3. Data Collection:**

* **Process:**
  + Generated synthetic data for 2500 users.
  + Included user demographics, financial transactions, and support interactions.
* **Data Cleaning:**
  + Filled missing values with median (numerical) or 'Unknown' (categorical).
  + Removed outliers using the IQR method.
* **Data Transformation:**
  + Added random noise and null values to simulate real-world imperfections.

**4. Data Analysis:**

* **Analytical Techniques:**
  + **EDA:** Histograms, bar charts, and pie charts for initial data exploration.
  + **Feature Engineering:** Created new features like income levels, total cashouts, and average tip amount.
* **Visualizations and Insights:**
  + **Gender Distribution:** Pie chart to show gender proportions.
  + **Region Distribution:** Bar chart to display user distribution across regions.
  + **Income Distribution:** Histogram to visualize income spread.
  + **Employer Distribution:** Bar chart to show employer frequency.
  + **Cash Out Frequency:** Histogram to analyze cash-out behavior.
  + **Tip Amount Distribution:** Bar chart to understand tip amounts.
  + **Zendesk Tickets:** Bar chart to evaluate support interactions.
  + **Lead Source Distribution:** Pie chart to show user acquisition channels.

**5. Implementation:**

* **Tools and Languages:**
  + **Python:** Data generation, preprocessing, and analysis.
  + **Pandas and NumPy:** Data manipulation and analysis.
  + **Power BI:** Visualization and dashboard creation.

**6. Evaluation:**

* **Performance Assessment**: The analysis successfully achieved the project objectives of understanding user demographics, income distribution, transaction behaviors, and support interactions. The visualizations provided clear insights into each aspect of the data, aligning well with the initial goals.
* **Challenges and Solutions**: During the analysis, challenges included handling missing values and outliers, as well as ensuring accurate data segmentation for feature engineering. These were addressed through rigorous data preprocessing steps, such as filling missing values with appropriate substitutes and removing outliers using the IQR method.
* **Validity and Reliability**: The results of the analysis are considered valid and reliable due to the thorough data cleaning and preprocessing steps taken. The use of various visualizations and segmentation techniques further ensured that the insights derived were accurate and meaningful.

**7. Conclusion:**

* **Key Findings:**
  + Gender and region distributions provide demographic insights.
  + Income distribution analysis highlights economic segments.
  + Cash-out and tip amount analyses reveal transaction behaviors.
  + Support interaction analysis helps identify user support needs.
* **Implications:**
  + Insights can inform marketing strategies, user support improvements, and financial planning.
  + Recommendations for future work include validating findings with real-world data and exploring additional features.

**8. References:**

* Python libraries: Pandas, NumPy
* Visualization tool: Power BI

**9. Appendices:**

* Code written to generate the date for the problem statement
  + 
* Data Cleaning
  + 