PhonePe Digital Payments Project

Task PhonePe Digital Payments Project.

Task 1: Data Loading and Understanding

1.1: Load each dataset and display its structure

1. Load the State_Txn and Users dataset and display its first 5 rows.

2. Load the State_TxnSplit dataset and display its bottom 10 rows.

3. Load the State_DeviceData dataset and display 10 rows from the middle of the dataset.

4. Load the District Txn and Users dataset and display its first 10 rows and last 10 rows.

5. Load the District Demographics dataset and display every 10th row.

1.2: Display basic statistics and data types for each dataset

1. For each dataset, display the summary statistics (mean, median, std, etc.) for numerical columns.

2. Display the data types of each column in each dataset.

1.3: Check for missing values

1. Identify any missing values in each dataset.

2. Calculate the percentage of missing values for each column that has missing values.

3. Highlight which column has the highest percentage of missing values in each dataset.

1.4: Create a summary

1. Calculate the total number of states and the total number of districts.

2. Identify the state with the highest number of districts.

Task 2: Exploratory Data Analysis (EDA)

2.1: Analyze transaction trends over the years for each state

2.2: Identify the most common transaction types in each state and quarter

1. For each state and quarter, determine the most frequent transaction type. Display the results in a tabular

format.

2.3: Determine the device brand with the highest number of registered users in each state

1. Identify the device brand with the highest number of registered users in each state. Display the results in a tabular format.

2.4: Create a list of the top district per state based on population

1. For each state, identify the district with the highest population. Display the results in a tabular format.

2. Create a column chart depicting the district with the highest population for each state.

2.5: Calculate the average transaction value (ATV) for each state

1. Compute the average transaction value for each state. Display the results in a tabular format.

2. Identify the top 5 states with the highest ATV and the top 5 states with the lowest ATV. Display the results.

2.6: Analyze app usage trends

1. Calculate the total number of app opens over the years and quarters for each state. Display the results in a tabular format.

2. Identify trends in app usage by creating a line plot showing the number of app opens over time for a selected state

2.7: Distribution of transaction types

1. Create a bar chart showing the distribution of different transaction types for each state for the most recent quarter in the dataset.

2.8: Find unique mapping between district name and district code

1. Identify the unique mapping between district names and district codes from the dataset.

2. Create a CSV file containing the unique district name and district code mappings.

3. Export the CSV file.

Task 3: Data Quality Checks

3.1: Ensure data consistency across state and district levels

1. For each state, calculate the total number of transactions, total transaction amount, and total registered users by summing up the values from the district level data.

2. Compare the results with the corresponding values at the state level to ensure they match.

3. Display any discrepancies found between the district-level and state-level data.

1. Calculate the total number of transactions and total transaction amount for each state over the years. Display the results in a tabular format.

2. Identify the top 5 states with the highest transaction volumes and the top 5 states with the lowest transaction volumes. Display the results.

Task 4: Data Merging and Advanced Analysis

4.1: Ratio of users to population by state

1. Merge the State_Txn and Users dataset with the District Demographics dataset to calculate the ratio of registered users to the population for each state. Display the results in a tabular format.

2. Create a column chart depicting the ratio of users to population by state.

4.2: Correlate population density with transaction volume

1. Merge the District Txn and Users dataset with the District Demographics dataset.

2. Calculate the correlation between population density and transaction volume.

3. Create a scatter plot to visualize the correlation between population density and transaction volume.

4.3: Average transaction amount per user

1. Merge relevant datasets to calculate the average transaction amount per user for each state. Display the results in a tabular format.

2. Identify the top 5 states with the highest average transaction amount per user and the top 5 states with the lowest average transaction amount per user. Display the results.

4.4: Device brand usage ratio

1. Merge the State_DeviceData dataset with the State_Txn and Users dataset.

2. Calculate the ratio of users using each device brand to the total number of registered users in each state. Display the results in a tabular format.

3. Create a bar chart depicting the device brand usage ratio for each state.

Task 5: Data Visualization

5.1: Plot the total transactions and amount over time for a selected state

1. Create a line plot showing the total number of transactions and the total transaction amount over time (years and quarters) for a selected state.

5.2: Create a pie chart showing the distribution of transaction types for a specific quarter

1. Create a pie chart showing the distribution of different transaction types for a selected state and quarter.

5.3: Visualize the population density of districts in a selected state

1. Create a bar plot showing the population density of districts in a selected state.

Task 6: Insights and Conclusions [Advanced Section]

6.1: Identify any trends or patterns in the transaction data

1. Analyze the transaction data to identify any noticeable trends or patterns. Summarize your findings.

6.2: Correlate demographic data with transaction data

1. Find correlations between demographic data (e.g., population density) and transaction data (e.g., transaction volume). Summarize your findings.

6.3: Summarize findings and insights

1. Summarize the key findings and insights from your analysis. Provide actionable recommendations based on the data.