EYouth
Data Analytics
Bootcamp

BANKING ANALYSIS
USING AI AND CLOUD TECHNOLOGIES

EYouth
Learning.com

❯ Start Slide

EYouth
Data Analytics
Bootcamp

Table of Contents

EYouth
Data Analytics
Bootcamp

Introduction

The Banking Analysis project leverages AI and cloud technology to enhance banking operations.

Powered by Azure, SQL Server, Power BI, and Python, this project delivers improved customer segmentation, robust risk management, and streamlined operational efficiency for smarter banking decisions.

PROJECT OVERVIEW

banks deal with massive data every day from customer transactions to loans and card activity. Without proper analysis, this data means nothing.

The Banking Analysis Dashboard helps solve this by:

- Tracking customer behavior and churn risk
- Monitoring account performance and balances
- Detecting anomalies in transactions
- Visualizing trends in loans, cards, and support cases

It turns raw banking data into clear, actionable insights.

EYouth
Data Analytics
Bootcamp

01 **What is the project ?**

The Banking Analysis Dashboard is a comprehensive data analytics solution built to transform raw banking data into meaningful insights.

It integrates SQL, Python, and visualization tools to monitor customer behavior, track account performance, and assess banking operations

02 **Why banking analytics**

The Banking Analysis Dashboard is a comprehensive data analytics solution built to transform raw banking data into meaningful insights.

It integrates SQL, Python, and visualization tools to monitor customer behavior, track account performance, and assess banking operations.

03 **Real-world use case**

Banks generate massive volumes of data every day. Without analytics, that data means nothing.

By applying data analytics, banks can reduce risk, improve customer retention, detect fraud, and optimize financial service

PROJECT GOALS

Data-Driven Decisions

Revolutionize banking with AI-powered customer insights.

Build Trust

Delight customers with fast, smart service.

Security & Stability

Stop fraud and predict loan risks instantly.

Dataset Structure

01 Customers

02 Accounts

03 Transactions

04 Loans

05 Credit Cards

06 Support Calls

EYouth
Data Analytics
Bootcamp

Technology Stack

01
Microsoft Azure
Cloud storage and processing of banking data.

02
SQL Server
Data management and querying for preprocessing.

03
Power BI
Visualization using Power Query, DAX, Measures, and Dashboards.

04
Python (Jupyter Notebook)
RFM segmentation, predictive modeling, and fraud detection.

05
AI Chatbot
AI-driven data interaction and customer query handling.

SYSTEM ARCHITECTURE

The system allows secure, real-time access to insights through multiple tools working on top of a single cloud database.

Azure → **Sql server** → **power ai** → **python notebook** → **ChatBot**

Azure: Hosts all cloud services including the SQL database

Sql server: Central storage for all banking data

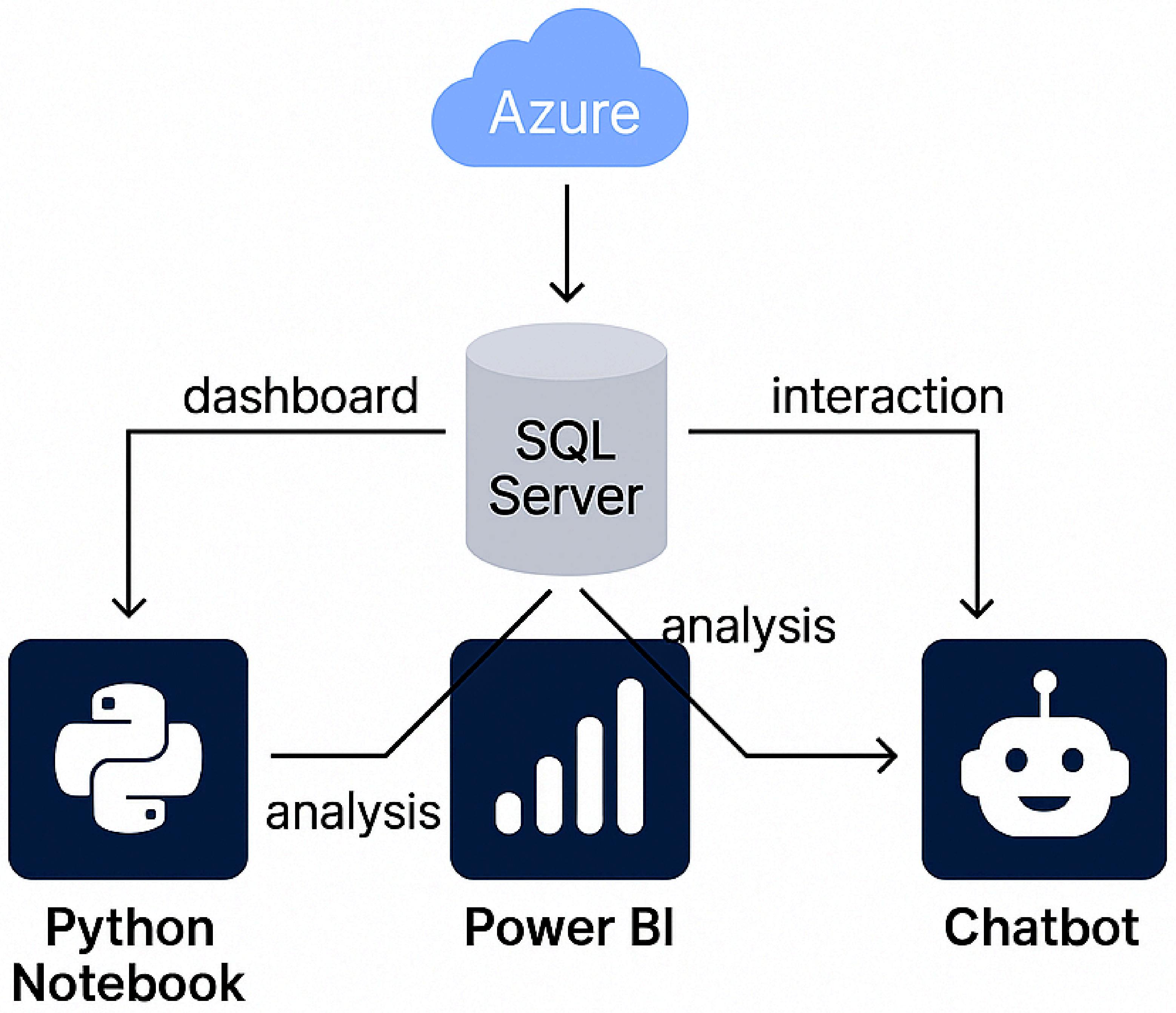
power ai: Connects to SQL to build interactive dashboards

python notebook: Pulls data from SQL for advanced analysis

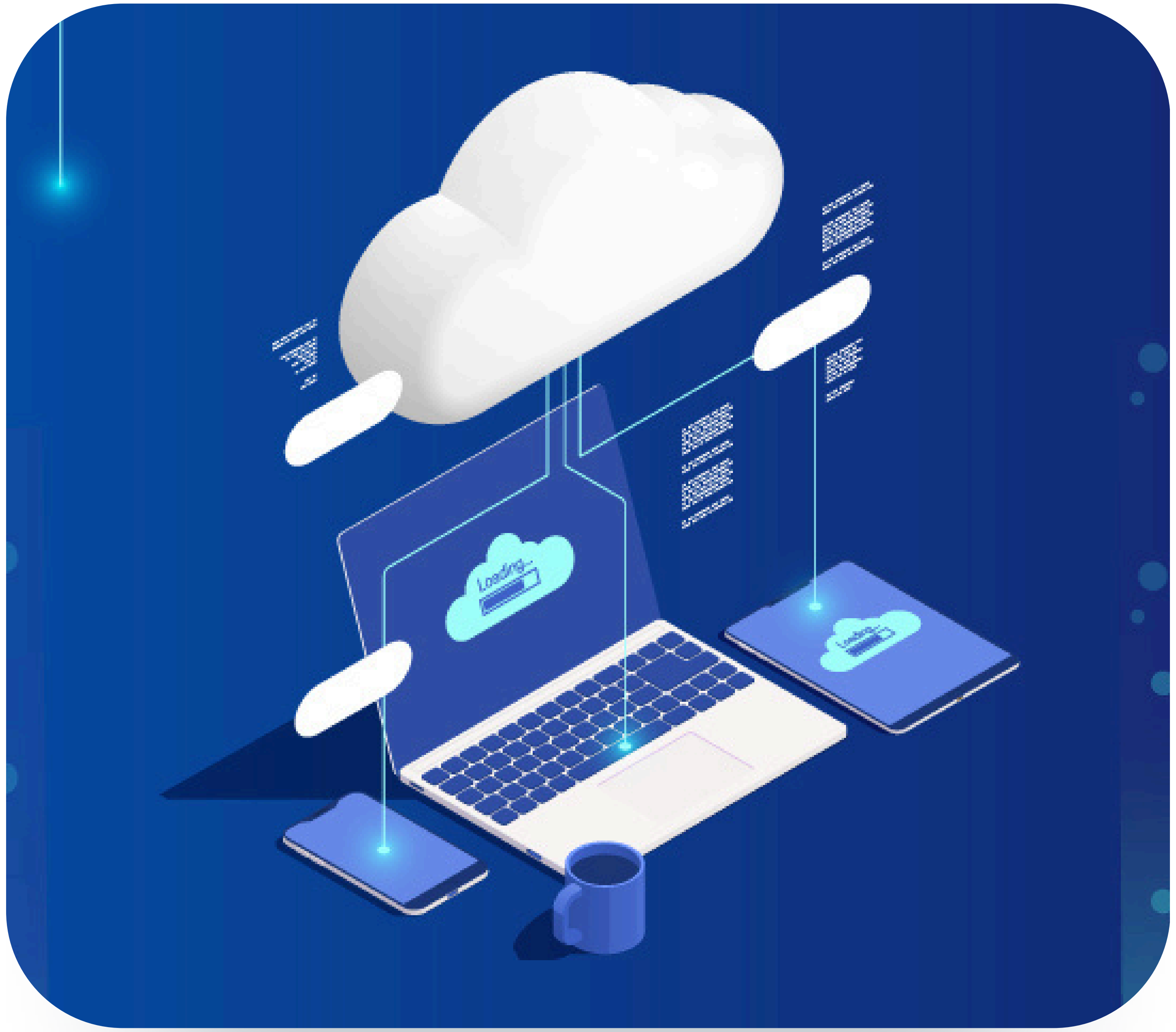
ChatBot: natural language queries and gets data via SQL

EYouth
Data Analytics
Bootcamp

System Architecture



EYouth
Data Analytics
Bootcamp



Microsoft Azure

Key Features and Benefits

It is used in this project only to create a SQL Server instance for hosting and managing the database to run SQL queries for customer analytics, account analysis, and fraud detection.

EYouth
Data Analytics
Bootcamp

AZURE SQL SERVER INTEGRATION

01

WHY CLOUD DB?

1. Accessible from anywhere by all tools (Power BI, Python, Chatbot)
2. Auto-scaling, backups, and disaster recovery built-in
3. Centralized source for all banking data
4. Secure access via Azure credentials and firewall settings

02

HOW IT WAS CONFIGURED

1. Created Azure SQL Server via Azure Portal
2. Uploaded SQL schema using SSMS / Query Editor
3. Defined database tables: Customers, Accounts, Transactions, etc.
4. Set firewall rules to allow trusted IPs
5. Generated secure connection strings for all tool

EYouth
Data Analytics
Bootcamp

SQL QUERIES &KPLS

CUSTOMER ANALYTICS

1. **Total Customers**

   Counts how many customers you have.

   Query:

```
SELECT COUNT(*) AS TotalCustomers
FROM Customers;
```

Results | Messages

	TotalCustomers
1	5000

2. **New Customers by Year**

   Shows how many new customers joined each year

   Query:

```
SELECT YEAR(JoinDate) AS JoinYear, COUNT(*) AS NewCustomers
FROM Customers
GROUP BY YEAR(JoinDate)
ORDER BY JoinYear;
```

Results | Messages

	JoinYear	NewCustomers
1	2015	346
2	2016	493
3	2017	488
4	2018	542
5	2019	509
6	2020	505
7	2021	504
8	2022	476
9	2023	506
10	2024	474
11	2025	157

5. **Churn Risks (no transactions in last 6 months)**

   Finds customers who have not done anything in 6 months, at risk of leaving.

   Query:

```
SELECT COUNT(DISTINCT c.CustomerID) AS Churn_Risk_Customers
FROM Customers c
LEFT JOIN Accounts a ON c.CustomerID = a.CustomerID
LEFT JOIN Transactions t ON a.AccountID = t.AccountID
WHERE t.TransactionID IS NULL
  OR t.TransactionDate < DATEADD(MONTH, -6, '2025-05-17');
```

	Churn_Risk_Customers
1	4960

6. **Customers with Accounts**

   Counts how many customers have accounts.

   Query:

```
SELECT COUNT(DISTINCT CustomerID) AS CustomersWithAccounts
FROM Accounts;
```

	CustomersWithAccounts
1	3171

TRANSACTION ANALYTICS

1. Total Transactions

Counts all transactions.

Query:

```
SELECT COUNT(*) AS TotalTransactions
FROM Transactions;
```

	TotalTransactions
1	20000

2. Transaction Type Distribution:

Shows how many transactions are deposits, withdrawals, etc.

Query:

```
SELECT TransactionType, COUNT(*) AS CountPerType
FROM Transactions
GROUP BY TransactionType;
```

	TransactionType	CountPerType
1	Payment	5056
2	Transfer	5055
3	Withdrawal	4919
4	Deposit	4970

3. Total and Average Amount by Transaction Type:

Adds up and averages the money for each transaction type.

Query:

```
SELECT
  TransactionType,
  COUNT(*) AS NumTransactions,
  SUM(Amount) AS TotalAmount,
  AVG(Amount) AS AvgAmount
FROM Transactions
GROUP BY TransactionType;
```

	TransactionType	NumTransactions	TotalAmount	AvgAmount
1	Payment	5056	25432001.4021139	5030.06356845607
2	Transfer	5055	25312679.808589	5007.45396806904
3	Withdrawal	4919	24384102.4509029	4957.12593025065
4	Deposit	4970	24980106.4753008	5026.17836525167

4. Transaction Analysis by Month:

Tracks how many transactions and how much money moved each month.

Query:

```
SELECT
  MONTH(TransactionDate) AS Month,
  FORMAT(TransactionDate, 'yyyy-MM') AS YearMonth,
  COUNT(*) AS TransactionsCount,
  SUM(Amount) AS TotalAmount
FROM Transactions
GROUP BY MONTH(TransactionDate), FORMAT(TransactionDate, 'yyyy-MM')
ORDER BY YearMonth ;
```

Top 5 Transactions by Amount:

Lists the 5 biggest transactions.

Query:

```
SELECT TOP 5 *
FROM Transactions
ORDER BY Amount DESC;
```

	TransactionID	AccountID	TransactionType	Amount	TransactionDate
1	7443	1463	Deposit	9999.8896484375	2025-04-04
2	17110	4432	Deposit	9999.4599609375	2023-09-26
3	6456	2207	Deposit	9999.25	2024-05-22
4	10101	3155	Deposit	9998.990234375	2024-07-07
5	3555	804	Deposit	9997.4697265625	2022-10-29

Average Transaction Value by Account Type:

Shows the average transaction size for each account type.

Query:

```
SELECT a.AccountType, AVG(t.Amount) AS Avg_Transaction_Value
FROM Transactions t
JOIN Accounts a ON t.AccountID = a.AccountID
GROUP BY a.AccountType;
```

	AccountType	Avg_Transaction_Value
1	Savings	5018.91191542289
2	Business	4981.94449105329
3	Checking	5016.78286683096

LOAN PORTFOLIO

Total Loans:

Counts all loans.

Query:

```sql
SELECT COUNT(*) AS TotalLoans
FROM Loans;
```

	TotalLoans
1	2500

Loan Type Distribution:

Shows how many loans are mortgages, personal loans, etc.

Query:

```sql
SELECT LoanType, COUNT(*) AS CountPerType
FROM Loans
GROUP BY LoanType;
```

	LoanType	CountPerType
1	Education	624
2	Personal	617
3	Car	633
4	Home	626

Total Loan Amount by Type:

Adds up the money lent for each loan type.

Query:

```sql
SELECT l.LoanType, SUM(l.LoanAmount) AS Total_Loan_Amount
FROM Loans l
GROUP BY l.LoanType;
```

	LoanType	Total_Loan_Amount
1	Education	154218864
2	Personal	154925690
3	Car	153848761
4	Home	153665578

Total and Average Loan Amount by Type:

Shows the number and average size of loans by type.

Query:

```sql
SELECT
  LoanType,
  COUNT(*) AS NumberOfLoans,
  SUM(LoanAmount) AS TotalLoanAmount,
  AVG(LoanAmount) AS AvgLoanAmount
FROM Loans
GROUP BY LoanType;
```

	LoanType	NumberOfLoans	TotalLoanAmount	AvgLoanAmount
1	Education	624	154218864	247145
2	Personal	617	154925690	251095
3	Car	633	153848761	243047
4	Home	626	153665578	245472

Average Interest Rate per Loan Type:

Shows the average interest rate for each loan type.

Query:

```sql
SELECT l.LoanType, AVG(l.InterestRate) AS Avg_Interest_Rate
FROM Loans l
GROUP BY l.LoanType;
```

	LoanType	Avg_Interest_Rate
1	Education	7.37951923219057
2	Personal	7.6685413240034
3	Car	7.36938388765706
4	Home	7.50738018313155

Loan Duration:

Shows how long each loan lasts in days.

Query:

```sql
SELECT
  LoanID,
  DATEDIFF(DAY, LoanStartDate, LoanEndDate) AS LoanDurationDays
FROM Loans;
```

	LoanID	LoanDurationDays
1	1	3308
2	2	1163
3	3	1196
4	4	1460
5	5	2531

Loan Issue by Year:

Counts loans given out each year.

Query:

```sql
SELECT
  YEAR(LoanStartDate) AS IssueYear,
  COUNT(*) AS LoansIssued
FROM Loans
GROUP BY YEAR(LoanStartDate)
ORDER BY IssueYear;
```

	IssueYear	LoansIssued
1	2020	371
2	2021	669
3	2022	636
4	2023	580
5	2024	244

CARD ISSUANCE & ACTIVITY

Total Cards:

Counts all issued cards.

Query:

```
SELECT COUNT(*) AS TotalCards
FROM Cards;
```

	TotalCards
1	4000

Card Type Distribution:

Shows how many cards are credit, debit, etc.

Query:

```
SELECT CardType, COUNT(*) AS CountPerType
FROM Cards
GROUP BY CardType;
```

	CardType	CountPerType
1	Prepaid	1355
2	Credit	1282
3	Debit	1363

Cards Expiring by Year:

Counts cards expiring each year.

Query:

```
SELECT
YEAR(ExpirationDate) AS ExpiryYear,
COUNT(*) AS CardsExpiring
FROM Cards
GROUP BY YEAR(ExpirationDate)
ORDER BY ExpiryYear;
```

	ExpiryYear	CardsExpiring
1	2026	584
2	2027	1044
3	2028	1002
4	2029	987
5	2030	383

Customer-Card Link:

Shows which customers have which cards.

Query:

```
SELECT
c.CustomerID,
c.FirstName + ' ' + c.LastName AS CustomerName,
ca.CardID,
ca.CardType,
ca.IssuedDate,
ca.ExpirationDate
FROM Cards ca
JOIN Customers c ON ca.CustomerID = c.CustomerID;
```

	CustomerID	CustomerName	CardID	CardType	IssuedDate	ExpirationDate
1	1	Dustin Diaz	564	Prepaid	2023-06-29	2027-10-23
2	1	Dustin Diaz	2932	Prepaid	2022-08-17	2027-08-13
3	5	Anna Bryant	1135	Prepaid	2023-09-01	2029-10-19
4	7	Andrew Watson	3114	Credit	2021-12-03	2029-05-02
5	8	Charles Leach	2571	Debit	2024-03-27	2026-10-02

Expired Cards:

Lists cards that expired before May 17, 2025.

Query:

```
SELECT *
FROM Cards
WHERE ExpirationDate < '2025-05-17';
```

	CardID	CustomerID	CardType	CardNumber	IssuedDate	ExpirationDate

Average Cards per Customer:

Shows the average number of cards per customer by card type

Query:

```
SELECT CardType, CAST(COUNT(CardID) AS FLOAT) / COUNT(DISTINCT
CustomerID) AS Avg_Cards_Per_Customer
FROM Cards
GROUP BY CardType;
```

	CardType	Avg_Cards_Per_Customer
1	Credit	1.1295154185022
2	Debit	1.13868003341688
3	Prepaid	1.15811965811966

SUPPORT PERFORMANCE

Total Number of Support Calls:

Counts all support calls.

Query:

```
SELECT COUNT(sc.CallID) AS Total_Support_Calls
FROM SupportCalls sc;
```

	Total_Support_Calls
1	3000

Resolved vs Unresolved Calls:

Shows how many calls were solved vs. still open.

Query:

```
SELECT
  COUNT(CASE WHEN sc.Resolved = 1 THEN sc.CallID END) AS Resolved_Calls,
  COUNT(CASE WHEN sc.Resolved = 0 THEN sc.CallID END) AS Unresolved_Calls
FROM SupportCalls sc;
```

	Resolved_Calls	Unresolved_Calls
1	1479	1521

Top Issue Categories:

Lists the most common reasons customers call.

Query:

```
SELECT sc.IssueType, COUNT(sc.CallID) AS Issue_Count
FROM SupportCalls sc
GROUP BY sc.IssueType
ORDER BY Issue_Count DESC;
```

	IssueType	Issue_Count
1	Transaction Dispute	774
2	Account Access	768
3	Loan Query	729
4	Card Issue	729

. **Loan Interest Rate Analysis:**

Shows average and range of interest rates for each loan type.

Query:

```
SELECT
  LoanType,
  AVG(InterestRate) AS AvgInterestRate,
  MAX(InterestRate) - MIN(InterestRate) AS InterestRateRange
FROM Loans
GROUP BY LoanType;
```

	LoanType	AvgInterestRate	InterestRateRange
1	Education	7.37951923219057	9.94999980926514
2	Personal	7.6685413240034	9.98000001907349
3	Car	7.36938388765706	9.95999979972839
4	Home	7.50738018313155	10

EYouth
Data Analytics
Bootcamp

Analysis of customer complaints and transactions

Identifying the number of transactions performed after a customer complaint was registered

	CustomerID	CustomerName	Phone	Email	ComplaintDate	TransactionsAfterComplaint	LastTransactionDate
1	1378	Donna Oneill	001-499-402-5961x51499	bakergeorge@white.net	2024-11-24	11	2025-05-05
2	213	William Murphy	001-846-230-7398x99456	collinmorales@douglas.net	2024-11-19	11	2025-03-11
3	2679	Kayla Chandler	787.529.8258x295	troy89@gmail.com	2025-01-28	10	2025-05-07
4	2672	Nicole Johnson	+1-183-489-5228x46953	patrickhorton@page-chen.com	2024-09-16	10	2025-04-06
5	4937	Joyce Savage	(570)862-7262	nhartman@hotmail.com	2024-08-27	10	2025-03-04
6	2823	Richard Bennett	9323620616	robertmorgan@thompson.com	2024-12-15	10	2025-02-06
7	1653	Sara Harris	001-507-857-4196x114	caseywilliam@yahoo.com	2024-06-14	9	2025-04-30
8	3550	Danny Young	001-602-996-3730x435	jenniferstrickland@yahoo.com	2024-08-06	9	2025-04-27
9	2464	Patricia Thomas	+1-263-152-0897x354	james79@baker-jackson.net	2024-07-25	9	2025-04-26
10	2842	Gilbert Todd	(762)970-3203x73544	cindy09@yahoo.com	2024-08-30	9	2025-01-24
11	1075	Sara Thompson	891-479-8056x53881	frojas@vaughn-walton.com	2024-10-26	8	2025-05-10
12	2364	Caleb Bennett	7440460662	toddwalter@rose.com	2024-09-05	8	2025-05-01
13	2377	James Schmidt	960-607-3013	guzmanbrian@gmail.com	2024-07-24	8	2025-04-18
14	79	Heather Barber	001-720-508-7478x760	oaguilar@yahoo.com	2024-05-18	8	2025-04-02
15	2302	Amanda Vance	556.116.5987x474	ramirezseth@yahoo.com	2024-08-27	8	2025-03-09
16	549	Jessica Cox	+1-002-802-5207	lewisclaudia@mendoza-alexander.com	2024-09-04	8	2025-02-07
17	3677	James Myers	(321)439-0915x59063	whiteandrew@lloyd.com	2024-11-11	7	2025-02-26
18	2243	Frank Hudson	088-297-1223	kellynathaniel@hotmail.com	2024-09-20	7	2024-12-20
19	2924	Sara Parker	(129)447-3147x33170	juanperry@gmail.com	2024-05-31	6	2025-05-07
20	402	John Wagner	(707)933-0426x57215	qford@brown-smith.biz	2025-01-05	6	2025-04-30
21	3670	Brenda Lopez	001-216-534-3208x9520	jennifer83@brown.com	2024-10-26	6	2025-04-27

EYouth
Data Analytics
Bootcamp

Comparison between number of transactions before and after the complaint

Analyzing trends to determine if there is an increase or decrease in transactions post-complaint

	CustomerID	CustomerName	CallDate	TransactionsBeforeComplaint	TransactionsAfterComplaint
1	319	Victor Johnson	2024-07-05	10	9
2	3401	Brittany Hampton	2024-08-06	12	7
3	213	William Murphy	2024-06-16	5	6
4	2231	Robert Lawson	2024-06-10	1	6
5	1264	Lauren Peterson	2024-06-10	11	6
6	226	Scott Walls	2024-07-30	12	6
7	4247	Emma Tucker	2024-09-09	2	6
8	3164	Amy Mitchell	2024-07-22	15	6
9	550	Gregory Peterson	2024-08-28	8	5
10	3437	Jimmy Burgess	2024-10-07	7	5
11	4245	Todd Vaughan	2024-07-28	11	5
12	3635	Kayla Smith	2024-06-18	9	5
13	1844	Katherine Johnson	2024-05-22	6	5
14	3480	Nancy Crawford	2024-10-27	6	5
15	1409	Brian Payne	2024-08-19	20	5
16	4583	Charles Thomas	2024-07-24	2	5
17	2764	Roger Smith	2024-08-04	7	5
18	1625	William Riddle	2024-06-26	6	5
19	3293	Kimberly Smith	2024-07-01	3	5
20	493	Emily Gonzales	2024-05-27	5	5
21	1870	Angela Austin	2024-09-21	15	5

First transaction after the complaint: How much Time taken to initiate the first transaction

Results | Messages

	CustomerID	CustomerName	IssueType	CallDate	FirstTransactionAfterCall	DaysBetweenCallAndFirstTransaction
1	3489	Andrea Ford	Transaction Dispute	2024-05-18	2025-04-23	340
2	4534	Jonathan Gomez	Card Issue	2024-05-30	2025-04-02	307
3	3254	Dennis Campbell	Transaction Dispute	2024-05-19	2025-03-14	299
4	245	Patrick Cook	Card Issue	2024-05-30	2025-03-24	298
5	1350	Christopher Richardson	Card Issue	2024-05-25	2025-03-18	297
6	3739	Michael Gardner	Card Issue	2024-05-11	2025-03-03	296
7	1096	Sandra Middleton	Card Issue	2024-05-23	2025-03-10	291
8	2094	Kara Love	Transaction Dispute	2024-06-19	2025-03-31	285
9	1432	Rebecca Arnold	Card Issue	2024-06-13	2025-03-21	281
10	10	Patricia Wilcox	Transaction Dispute	2024-06-14	2025-03-20	279
11	3001	Joanna Finley	Transaction Dispute	2024-07-30	2025-05-04	278
12	245	Patrick Cook	Card Issue	2024-06-21	2025-03-24	276
13	3409	Jodi Smith	Card Issue	2024-07-11	2025-04-11	274
14	2349	Marie Lester	Transaction Dispute	2024-07-20	2025-04-18	272
15	3077	Kim Sullivan	Card Issue	2024-06-08	2025-03-05	270
16	3737	Deborah Carroll	Transaction Dispute	2024-06-08	2025-03-02	267
17	3523	Nicole Mckenzie	Transaction Dispute	2024-06-16	2025-03-08	265
18	2697	Eric Lucas	Card Issue	2024-06-01	2025-02-20	264
19	3420	Tyler Mathews	Transaction Dispute	2024-07-19	2025-04-08	263
20	3051	Barbara Jackson	Card Issue	2024-06-01	2025-02-17	261
21	1446	Danielle Vaughn	Transaction Dispute	2024-07-02	2025-03-20	261

EYouth
Data Analytics
Bootcamp

POWER BI

POWER BI INTEGRATION

EYouth

01

Why Power BI

- Real-time connection to Azure SQL
- Clean, interactive dashboards for KPIs
- Drag-and-dropvisualizations for fast insights
- Ideal for non-technical stakeholdere

02

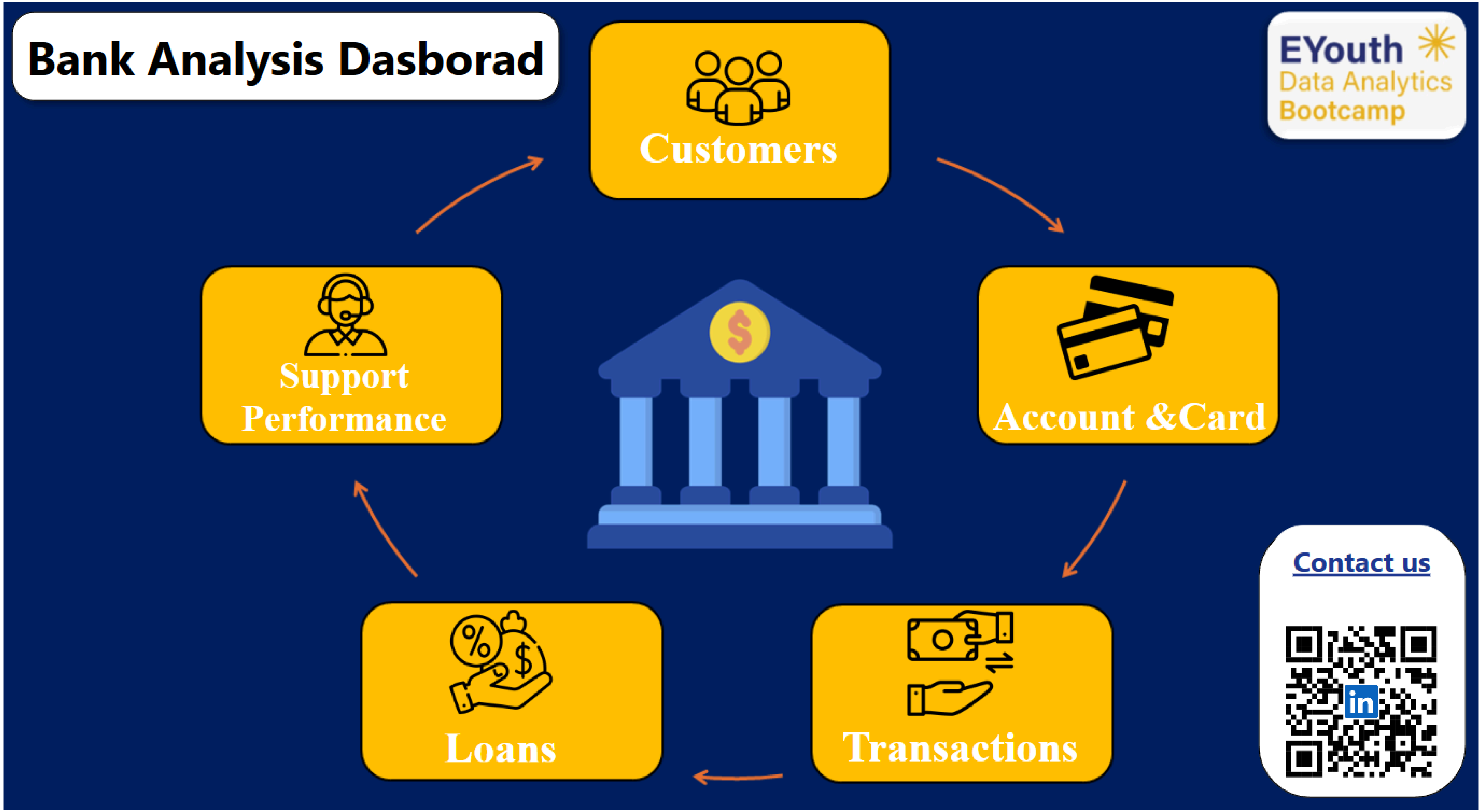
Dashboards Built

- Customer analysis
- Accounts & cards analysis
- Transactions analysis
- Loans analysis
- Support performance analysis

03

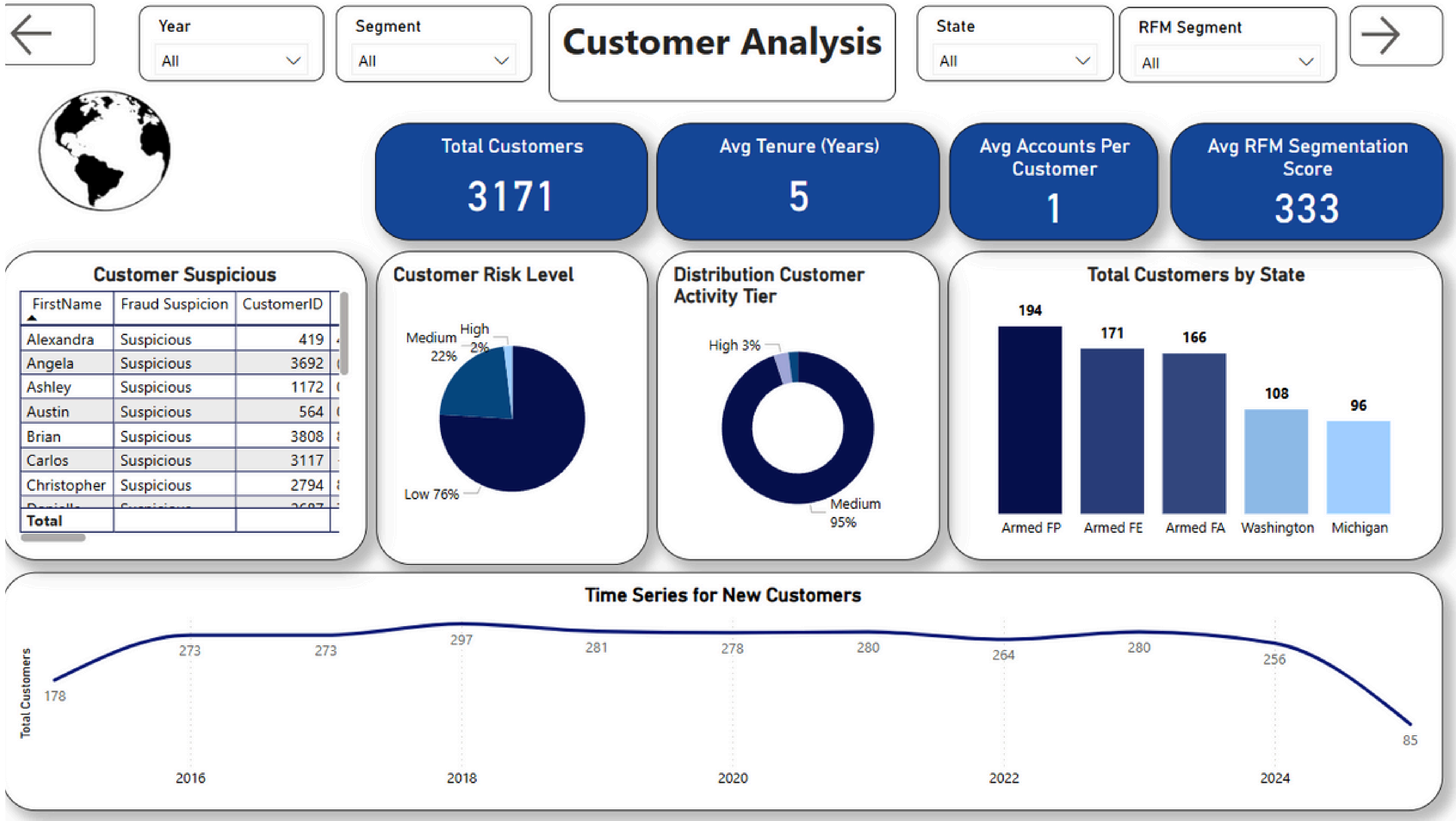
How It Was Connected

Ultimately, our goal is to drive tangible results for your business. We focus on optimizing conversion rates through strategic marketing initiatives. Whether it's through compelling calls-to-action, or optimized user experiences.


Bank Analysis Dasborad
Customers
Account &Card
Transactions
Loans
Support Performance
EYouth
Data Analytics Bootcamp
Contact us
in

POWER BI

EYouth



Key Metrics

- Total Customers: 3,171
- Avg Tenure: 5 years
- Avg Accounts Per Customer: 1
- Risk Segments:
    - Low: 76%
    - Medium: 22%
    - High: 2%

Insights

Most customers are low-risk and have average activity. Suspicious behavior is flagged early using segmentation. This allows proactive fraud detection and targeted customer retention strategies

POWER BI

EYouth

Map Overview

- Strong concentration in southern and central US states
- Sparse but strategic presence in Europe (Italy, Spain)
- Coverage includes Hawaii and parts of Northwest US

Use Cases

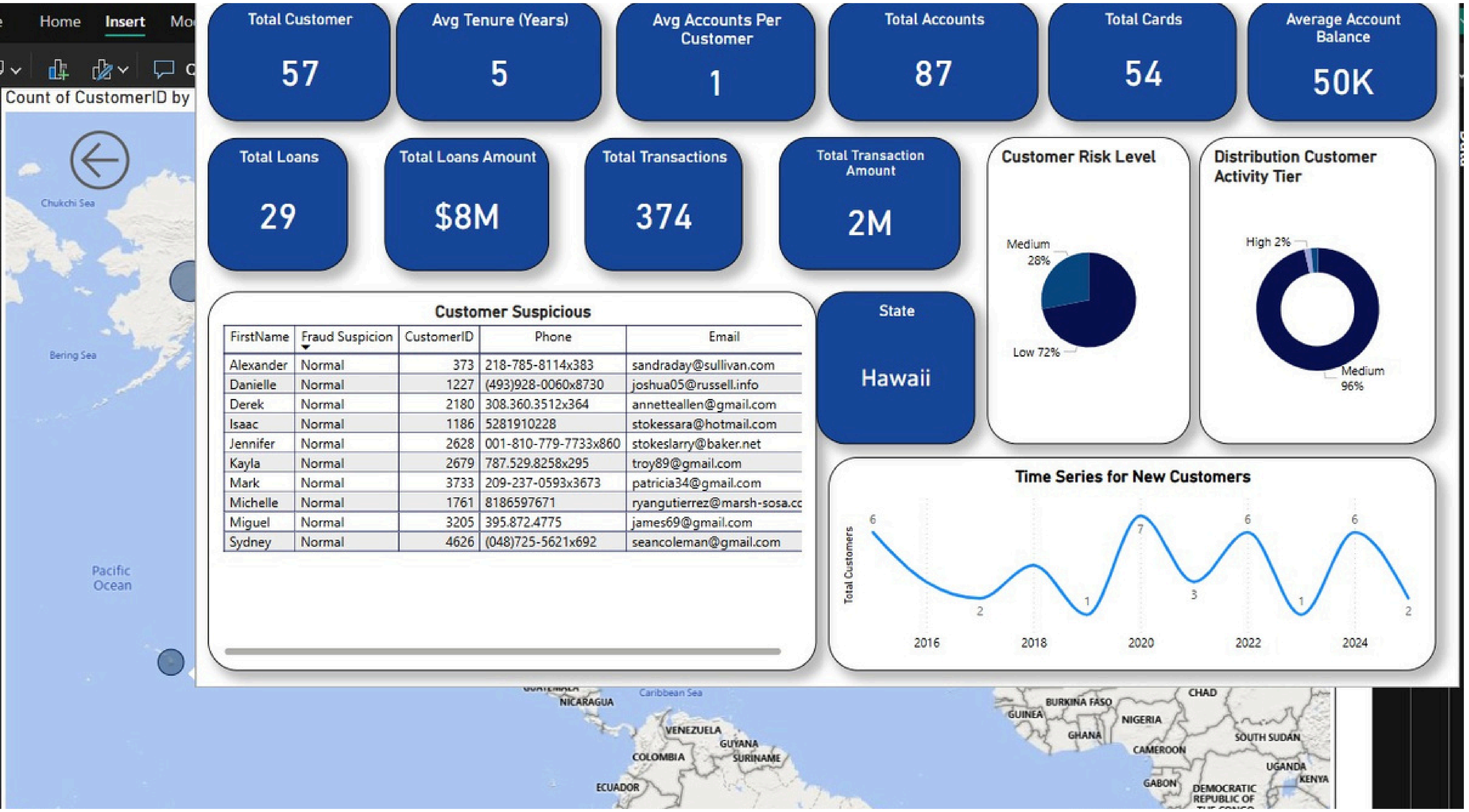
- Region-wise segmentation of customer base
- Helps prioritize marketing efforts and risk monitoring
- Supports decisions on branch expansion or localized campaigns

Insights

The geographic heatmap helps highlight customer density by state, detect growth zones, and monitor location-based risks.

POWER BI

EYouth



Key Metrics

- Total Customers: 57
- Total Loans: 29 | Loan Amount: $8M
- Total Transactions: 374 | Transaction Value: $2M
- Total Accounts: 87 | Cards: 54
- Average Account Balance: $50k

Segmentation & Risk

- Customer Risk:
- Low: 72%
- Medium: 28%
- Activity Tier:
- Medium: 96%
- High: 2%

Use Cases

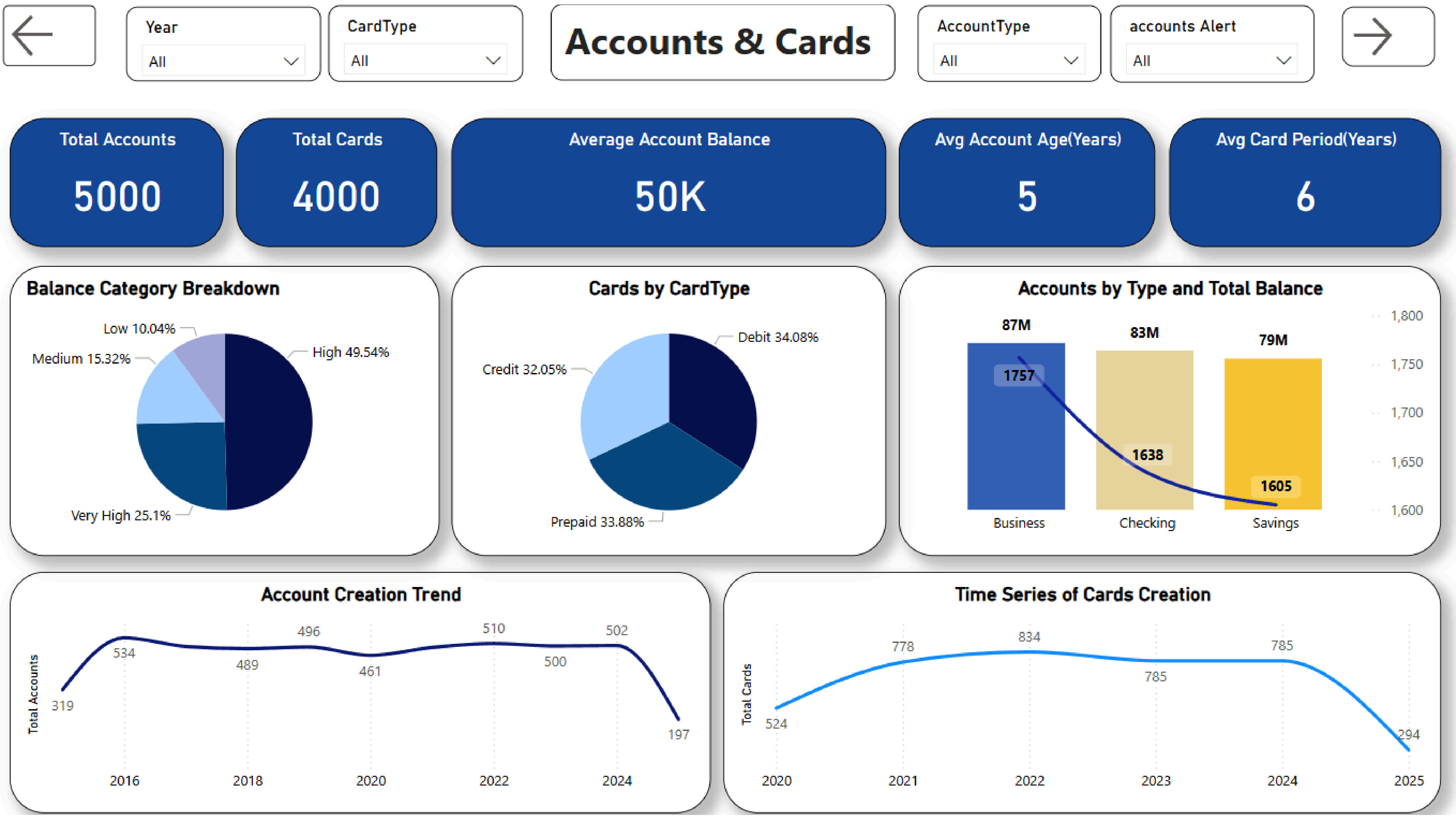
Snapshot of customer behavior and risk levels.

Most users are low-risk, but new customer growth is inconsistent.

Suspicious profiles flagged early using fraud detection logic

POWER BI

EYouth



Key Metrics

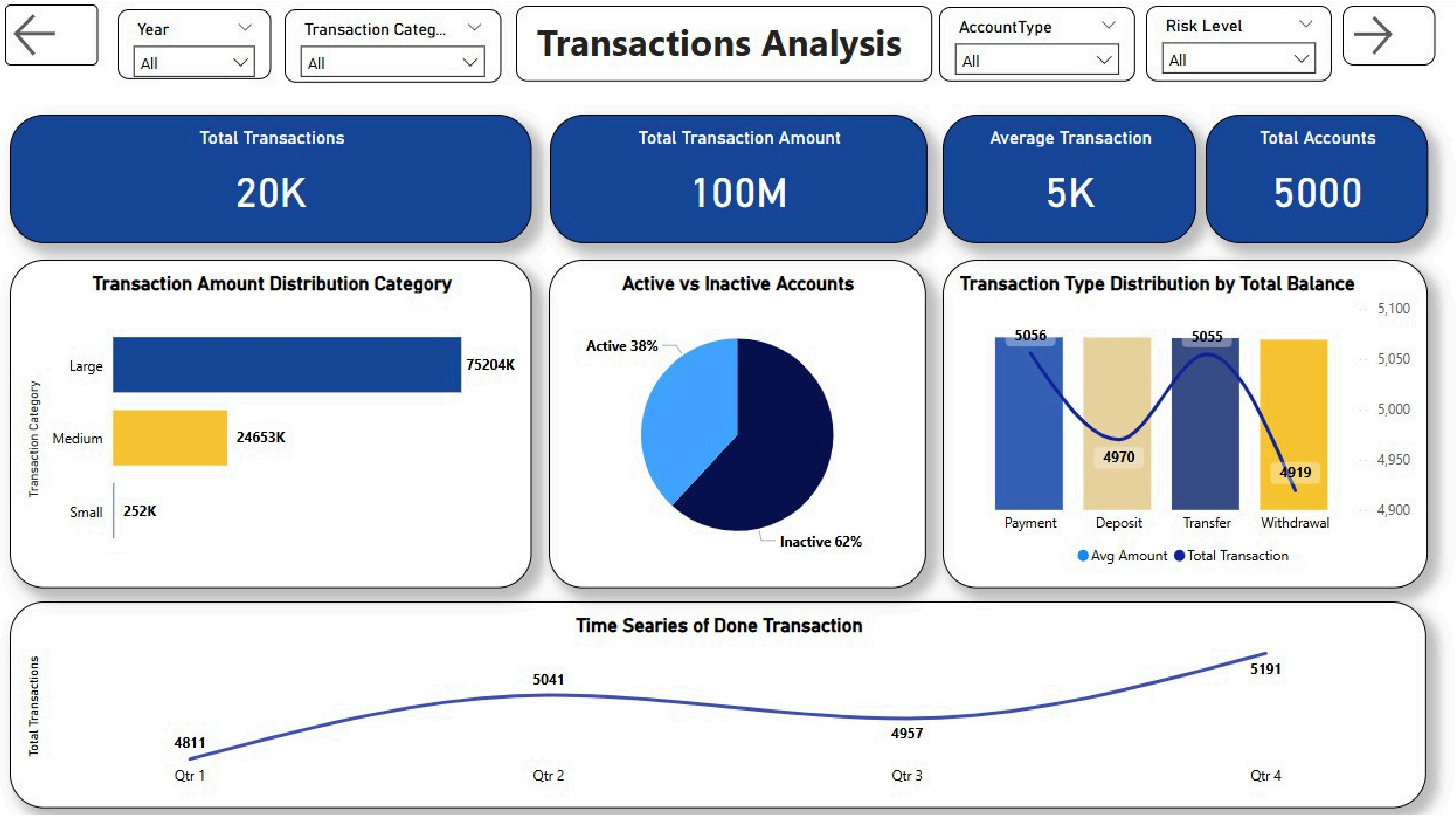
- Total Accounts: 5,000
- Total Cards: 4,000
- Avg Account Balance: $50K
- Cards by Type:
- Prepaid: 34%
- Credit: 32%
- Debit: 34%

Insights

Business accounts hold the highest total balance. Prepaid card usage is strong, offering growth potential for loyalty programs.

POWER BI

EYouth



Key Metrics

- Total Transactions: 20,000
- Total Transaction Amount: $100M
- Average Transaction: $5K
- Accounts: 5,000
- Active vs Inactive:
- Active: 38%
- Inactive: 62%

Transaction Types

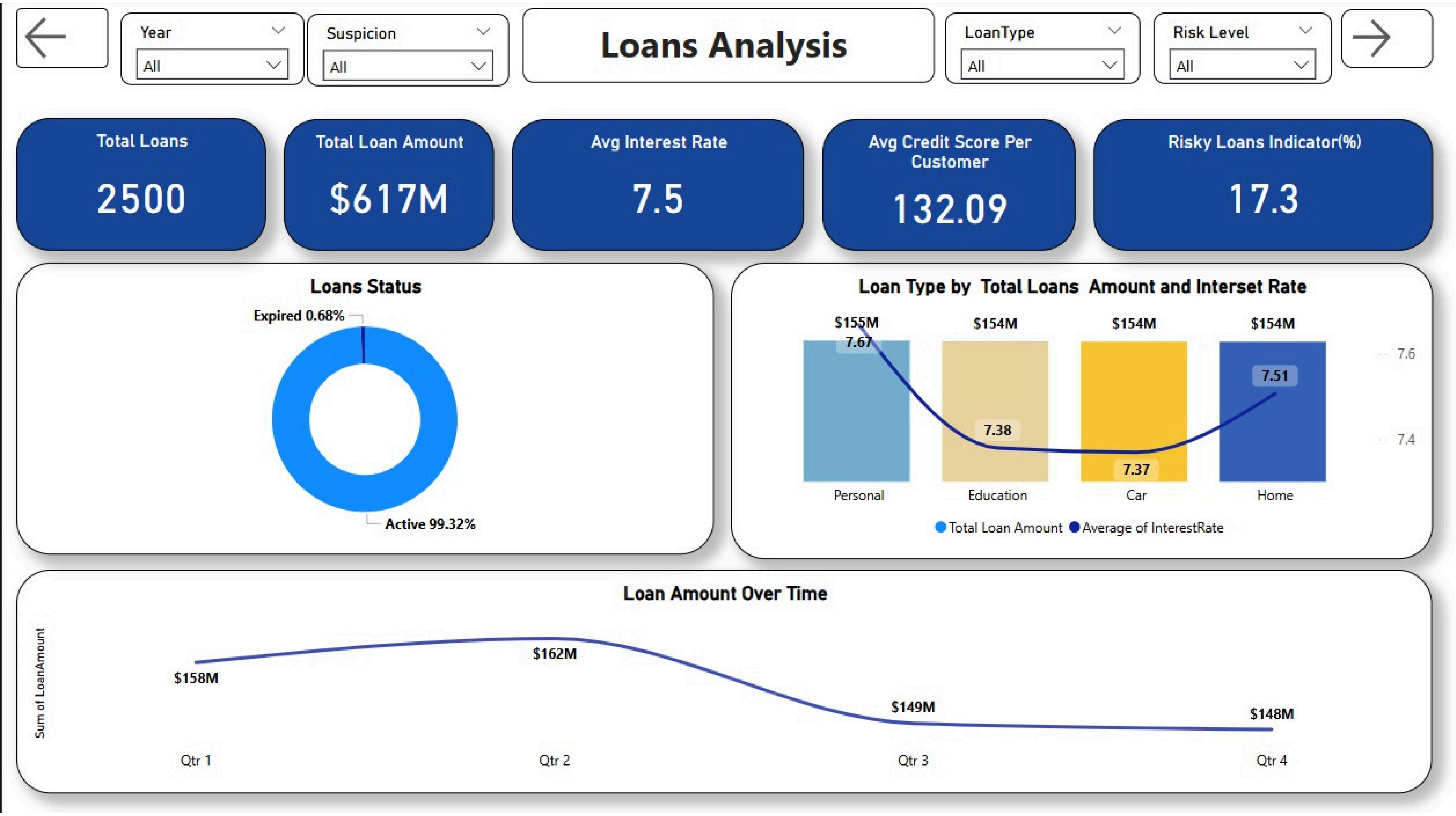
- Payment: 5,056
- Deposit: 4,970
- Transfer: 5,055
- Withdrawal: 4,919
- Most value: Large transactions ($75M

Insights

High volume is concentrated in a small number of large-value transactions. Many accounts are inactive, indicating opportunity to re-engage customer

POWER BI

EYouth



Key Metrics

- Total Loans: 2,500
- Total Loan Amount: $617M
- Avg Interest Rate: 7.5%
- Avg Credit Score Per Customer: 132.09
- Risky Loans: 17.3%
- Loan Status:

  Active:99.32%

  Expired:0.68%

Loan Distribution

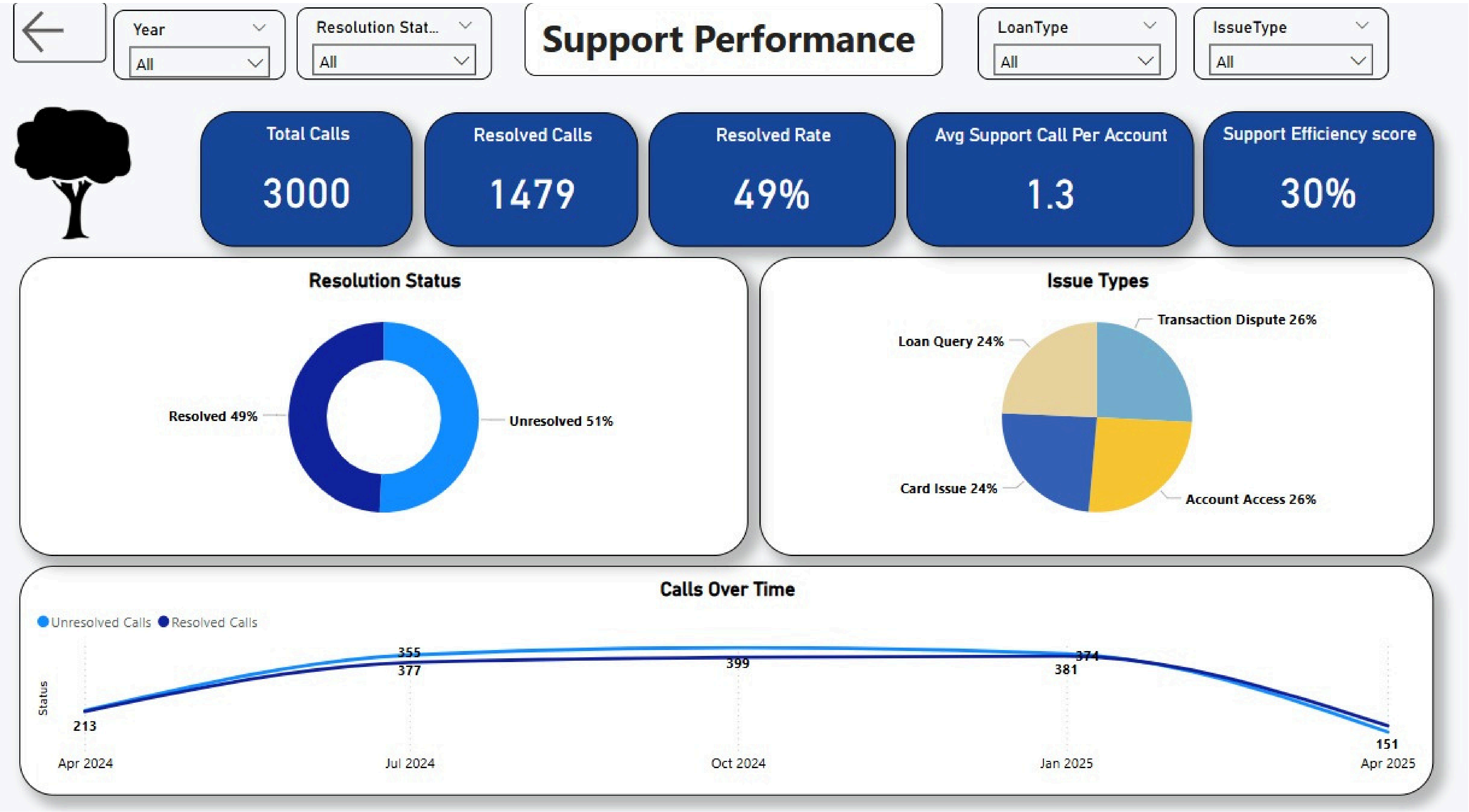
- Personal: $155M @ 7.67%
- Education: $154M @ 7.38%
- Car: $154M @ 7.37%
- Home: $154M @ 7.51%

Insights

Although the portfolio is well distributed, personal loans show slightly higher interest rates, suggesting more risk. Monitoring needed for high-risk borrowers

POWER BI

EYouth

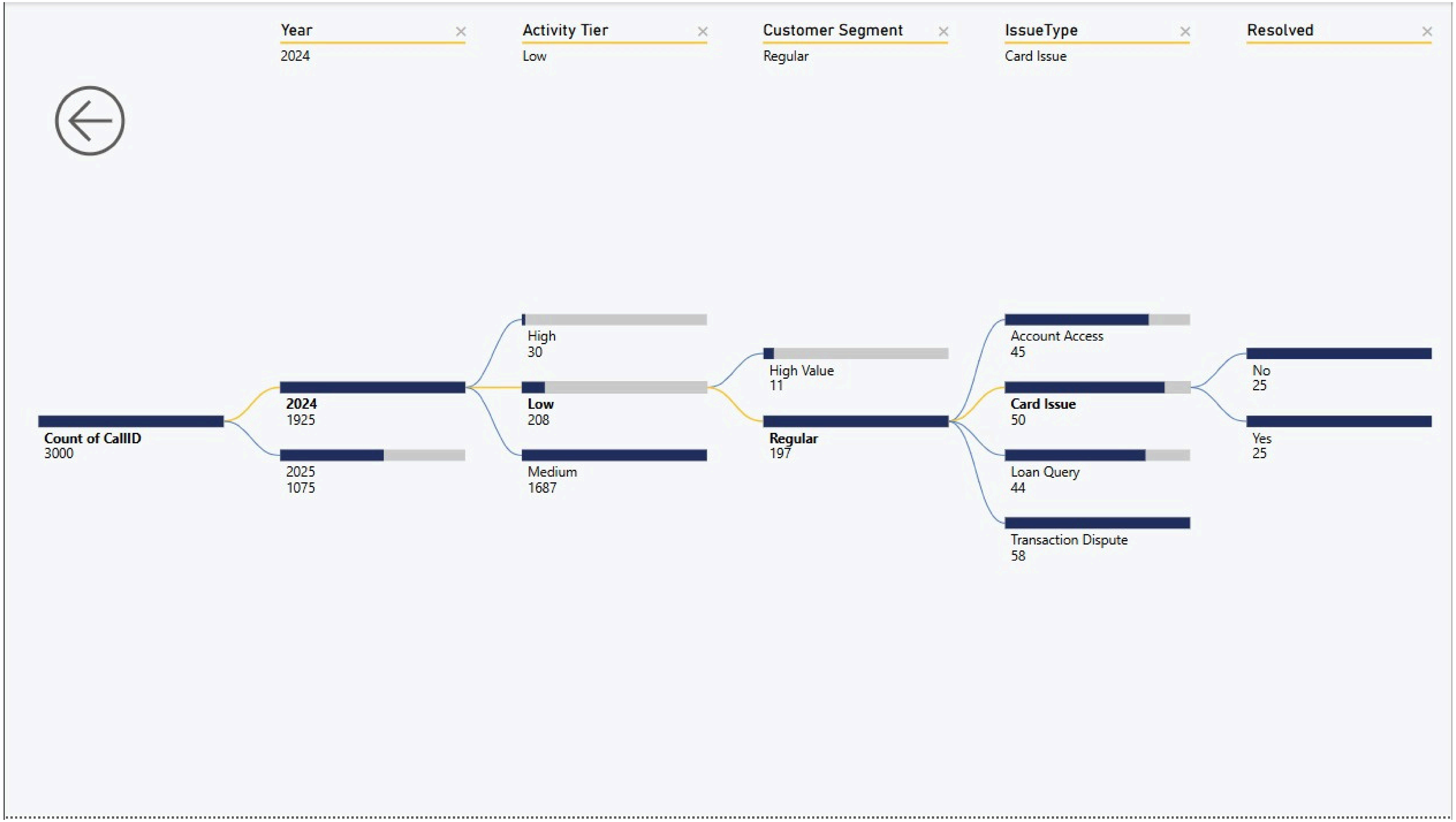


Key Metrics

- Total Calls: 3,000
- Resolved: 1,479 (49%)
- Avg Calls per Account: 1.3
- Support Efficiency: 30%
- Most Common Issues:
- Card Issues
- Transaction Disputes
- Loan Queries

Insights

Only half the support cases are resolved, indicating a need to improve support operations. Card issues are frequent, requiring better onboarding or clarity


Year
2024
Activity Tier
Low
Customer Segment
Regular
IssueType
Card Issue
Resolved
Count of CallID
3000
2024
1925
2025
1075
High
30
Low
208
Medium
1687
High Value
11
Regular
197
Account Access
45
Card Issue
50
Loan Query
44
Transaction Dispute
58
No
25
Yes
25

EYouth
Data Analytics
Bootcamp

Python Notebook

Why Python Notebooks?

Python notebooks (Jupyter) were used to go beyond SQL and perform deeper, custom analytics. They allowed flexibility in building machine learning models, creating complex plots, and exploring data iteratively

Notebook Goals

- Clean and transform raw banking data
- Perform advanced statistical analysis
- Detect anomalies and suspicious behaviors
- Segment customers into behavior-based clusters
- Forecast future values (loans, transactions

Tools & Libraries Used

- pandas, numpy – Data handling
- seaborn, plotly – Visualization
- scikit-learn – Classification, clustering, anomaly detection
- pmdarima, SARIMAX – Time series forecasting

Python Jupyter Notebook

RFM Customer Segmentation

Identifies Best , Loyal and At-Risk customers to optimize marketing strategies

Risk Prediction

Uses Random Forest to predict loan risks, enhancing financial security

Fraud Detection

Combines rule-based and AI-based methods to flag suspicious transactions

Forecasting

Predicts transaction over time and loan trends for strategic planning

Python Libraries

Chatbot & Notebook

1.Data Manipulation

- **pandas:** Handle structured data (DataFrames)
- **numpy:** Perform mathematical operations on arrays

2.Data Visualization

- **matplotlib.pyplot:** Build static charts
- **seaborn:** Generate statistical plots easily
- **plotly.express / graph_objects:** Create interactive dashboard

3. Date & Time Handling

- **datetime, timedelta:** Format and calculate time intervals

4. Machine Learning

- **train_test_split:** Split data for training/testing
- **RandomForestClassifier:** Predict risk levels
- **classification_report, confusion_matrix:** Evaluate model performance
- **IsolationForest:** Detect fraud/anomalies
- **StandardScaler:** Normalize features
- **KMeans:** Cluster customers

5. Time Series Analysis

- **adfuller:** Check stationarity
- **auto_arima:** Auto model selection
- **SARIMAX:** Forecast seasonal trends (loans, transactions)

6. Core ML Framework

- **scikit-learn (sklearn):** Unified machine learning library used across all models

DATA PREPARATION

What was done:

- Loaded Excel file containing 6 datasets:
- Customers, Accounts, Transactions, Loans, Cards, SupportCalls
- Checked and handled missing values in each sheet
- Converted all relevant date columns to datetime format
- Exported a CSV summary of missing data for validati

- Data cleaning was essential to avoid analysis errors, especially with time-based analytics.
- Ensuring consistency across all datasets allowed seamless merging, grouping, and time series modeling

MACHINE LEARNING MODELS

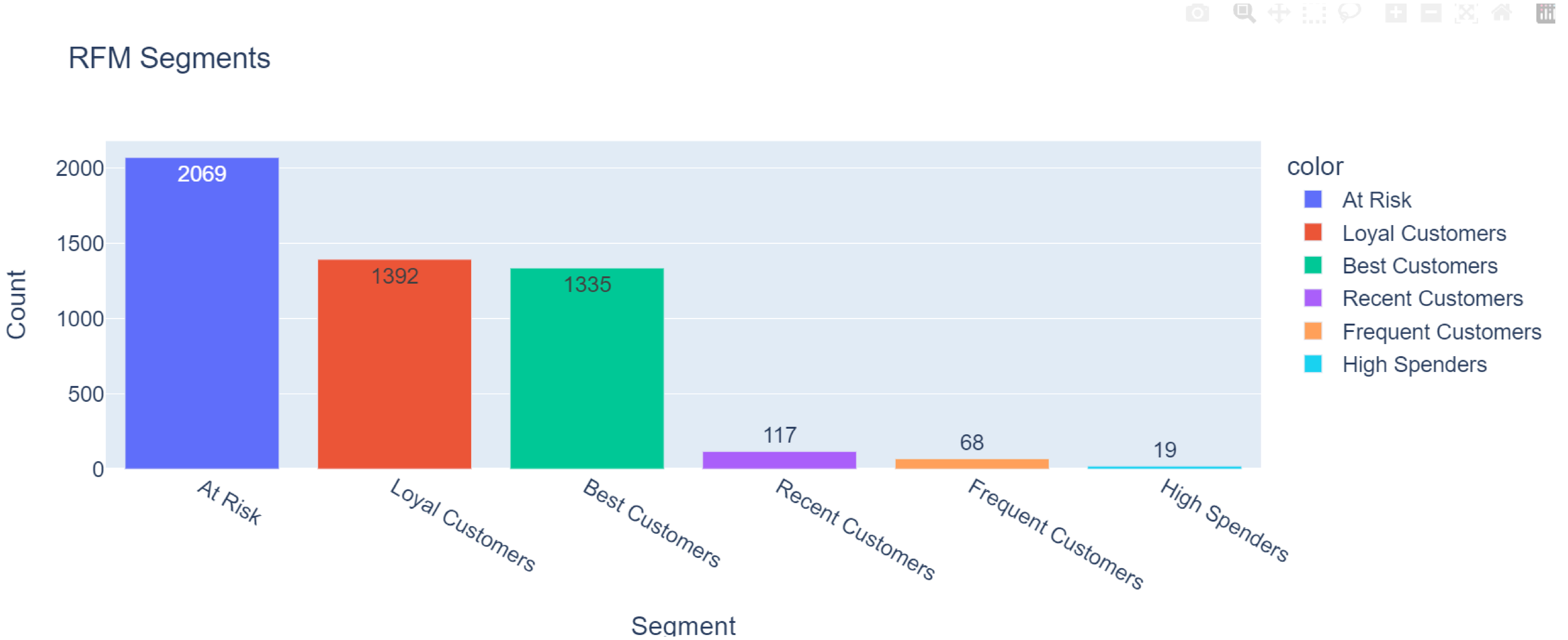
Models Used

- 1. Random Forest Classifier (Supervised)
- Used for predicting customer risk levels
- Input features: account balance, activity, tenure, card type
- Evaluated using confusion matrix & classification report

- 2. Isolation Forest (Unsupervised)
- Identified anomalies and suspicious behaviors
- Flagged outliers based on abnormal transaction patterns and balance spikes

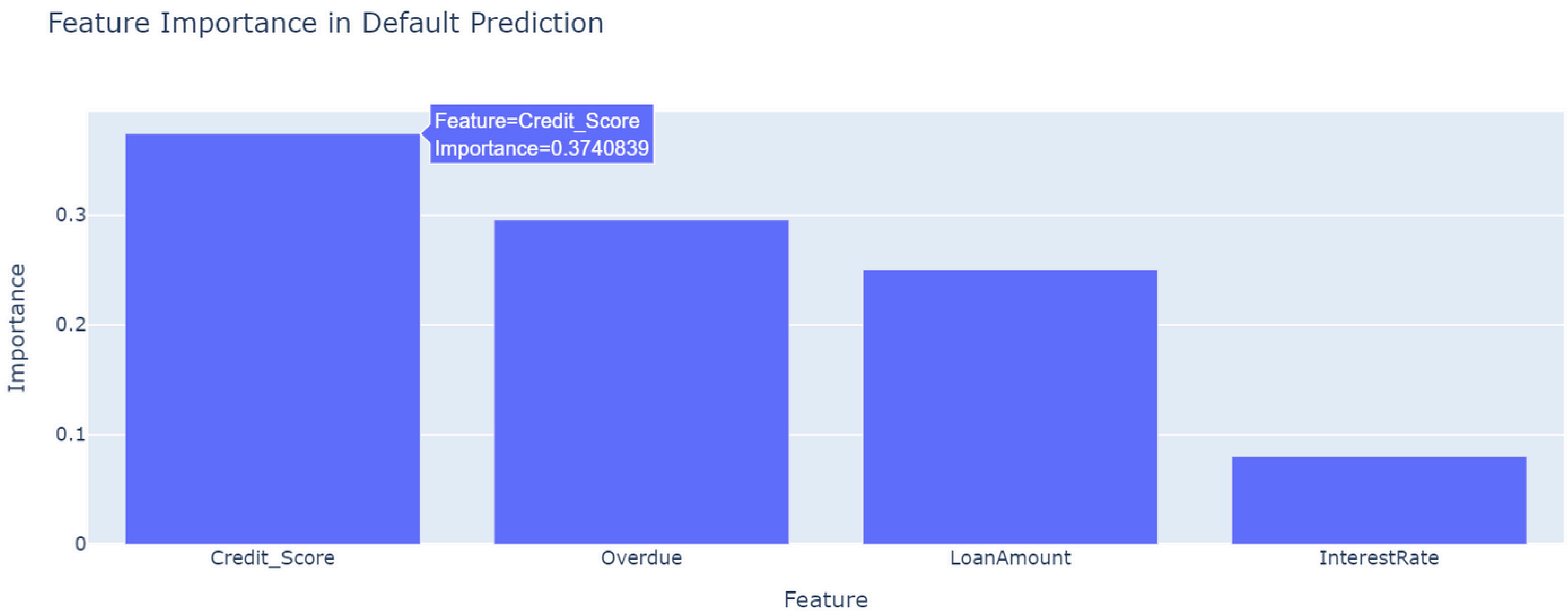
- 3. KMeans Clustering (Unsupervised)
- Segmented customers into 3 behavioral clusters
- Based on similarity in balance, tenure, and activity
- Visualized as colored scatter plots (cluster separation)

RFM CUSTOMER SEGMENTATION



RISK PREDICTION

EYouth



predict customer risk based on...

- Credit Score
- Overdue
- Loan Amount
- Interest Rate

Feature Importance

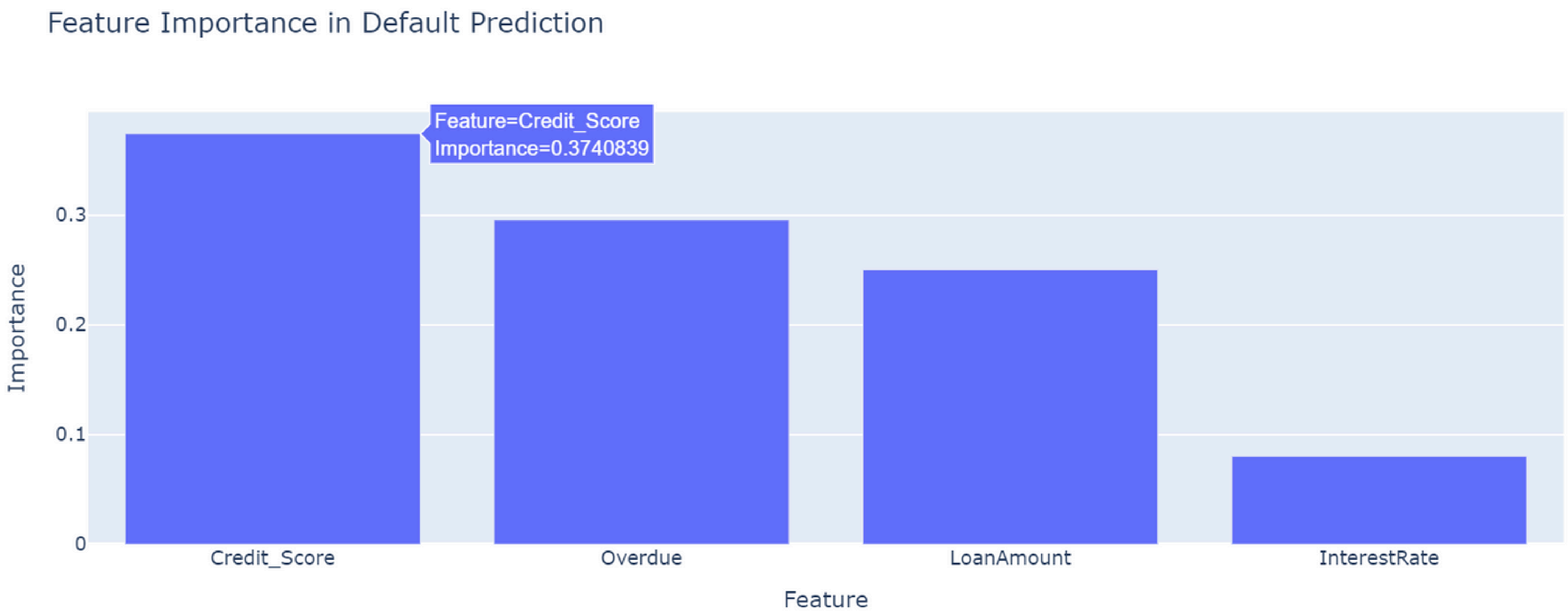
- Credit_Score: 37%
- Overdue: The second strongest indicator
- LoanAmount و InterestRate: less effected

Insight

Credit score and overdue behavior are the top risk indicators.
This helps the bank prioritize follow-up actions for high-risk profile

RISK PREDICTION

EYouth



predict customer risk based on...

- Credit Score
- Overdue
- Loan Amount
- Interest Rate

Feature Importance

- Credit_Score: 37%
- Overdue: The second strongest indicator
- LoanAmount و InterestRate: less effected

Insight

Credit score and overdue behavior are the top risk indicators.
This helps the bank prioritize follow-up actions for high-risk profile

TIME SERIES FORECASTING

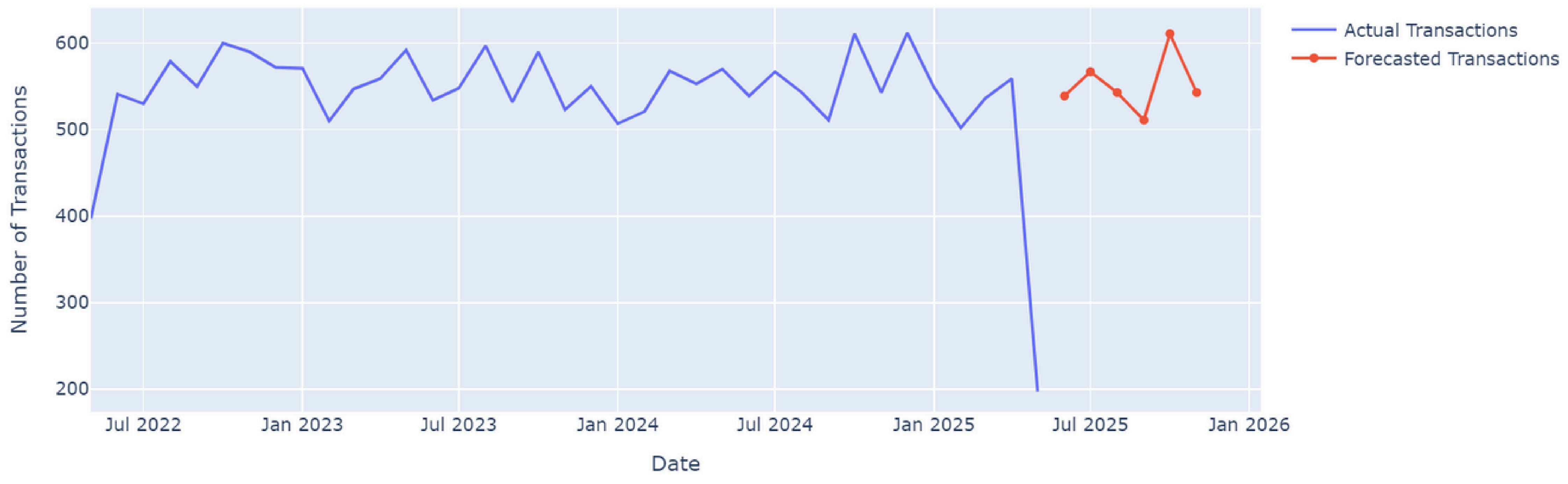
Models Applied

1. ADF Test (adfuller)
   - Checked stationarity of the time series
   - Ensured the data was suitable for modeling (no trend/seasonality distortion)

2. auto_arima
   - Automatically determined optimal ARIMA model parameters
   - Reduced trial-and-error in model tuning

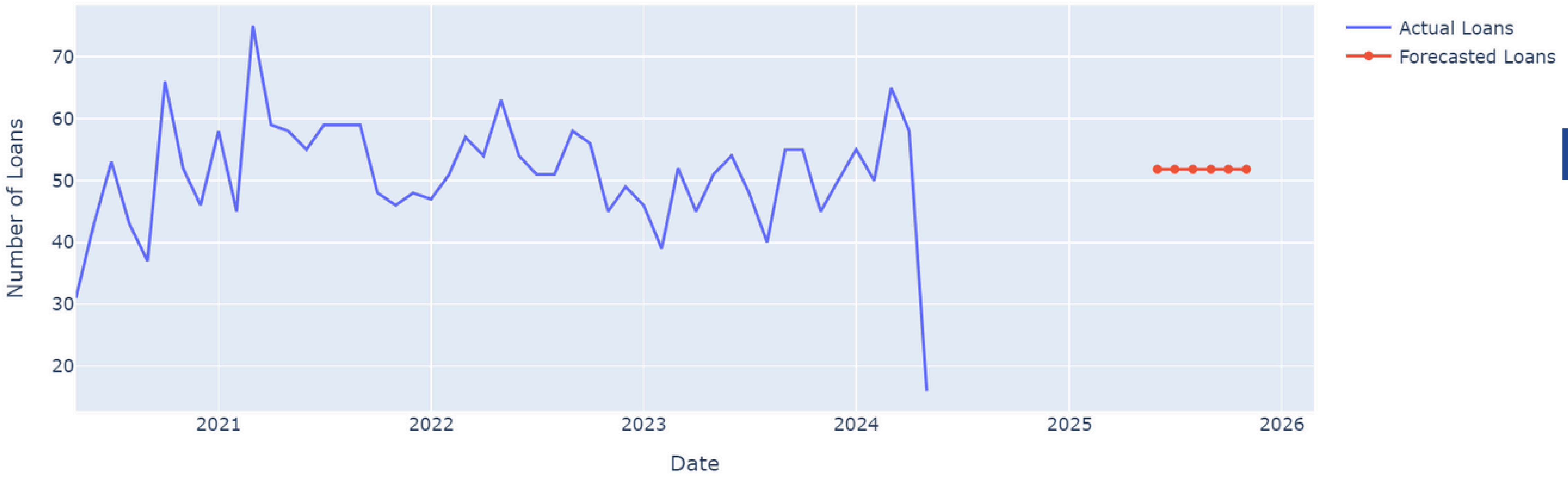
3. SARIMAX Model
   - Used for forecasting loan trends and transaction volumes
   - Captured seasonality and external regressors if needed

FORECASTING – TIME SERIE



Transaction Forecast

- Shows expected increase in transactions by mid-2025
- Used auto_arima for modeling + validation

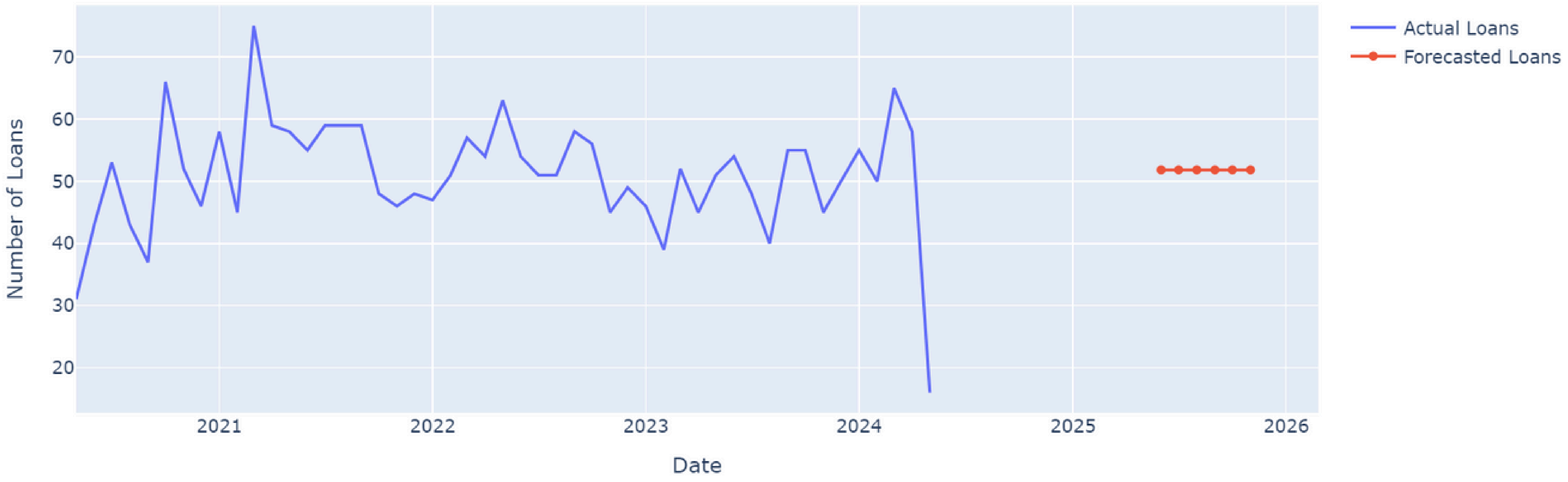


Loan Forecast :

- Based on SARIMAX model
- Shows loan trend from 2021 to 2025
- Forecast indicates stability with slight decline post-2024

FORECASTING – TIME SERIE

EYouth



Transaction Forecast

- Shows expected increase in transactions by mid-2025 (upper & lower)
- Used auto_arima for modeling + validation

Loan Forecast :

- Based on SARIMAX model
- Shows loan trend from 2021 to 2025
- Forecast indicates stability with slight decline post-2024

AI CHATBOT

EYouth
Data Analytics
Bootcamp

Chatbot Integration

Purpose:

Enable users (even non-technical) to interact with banking data using natural language — no need to write queries or use dashboards manually.

01. How It Works

- Trained on the same banking dataset used in Power BI & notebook
- Uses keyword extraction or LLM prompts to understand the question
- Runs SQL or filtered Pandas queries under the hoo

02. Insight

The chatbot brings your data system to life by turning static datasets into a conversational assistant.
It saves time, removes technical barriers, and enables any stakeholder to extract insights in secon

app.py:

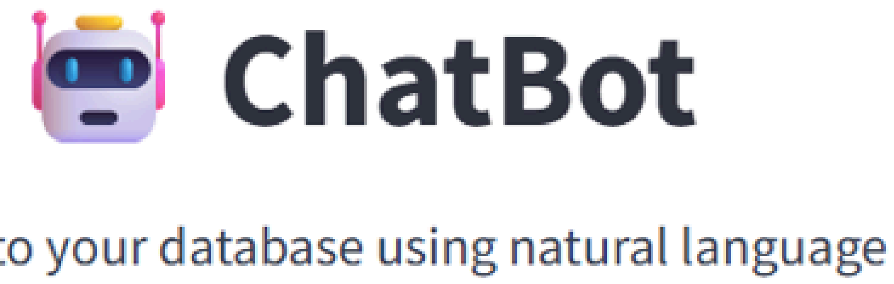
This is the Streamlit file; it handles the user interface.

agent.py:

Contains the chatbot logic. It receives the user's question, builds the conversation history, interacts with the language model, etc.

logging.py:

Handles structured and colored logging in the terminal using the rich library.



models.py:

Defines how to load and configure the language model (Ollama or OpenRouter) with the appropriate settings.

config.py:

Holds the project configurations, such as the model name, model type, database path.

tools.py:

Contains the tools that the model can use, such as list_tables, sample_table, describe_table, and execute_sql.

EYouth
Data Analytics
Bootcamp

File: data/banking.sqlite

Size: 3.18 MB

Tables:

- Banking_Transactions (20000 rows)
- Banking_Loans (2500 rows)
- Banking_Cards (4000 rows)
- Banking_SupportCalls (3000 rows)
- HR_Employees (1000 rows)
- HR_Hiring (1000 rows)
- HR_SickLeaves (2500 rows)

ChatBot

Talk to your database using natural language

Created by **EYouth Bootcamp - TEAM 20**

What's The Number of Transactions in 2022

The number of transactions in 2022 is 4,359.

What is the average amount of all transactions?

The average transaction amount is approximately 5005.44 USD.

Type your message...

EYouth
Data Analytics
Bootcamp

Database Information

File: data\banking.sqlite

Size: 3.18 MB

Tables:

Banking_Transactions (20000 rows)

	Amount	TransactionDate
0	3150.12	2023-09-24 00:00:00
1	6212.12	2022-06-07 00:00:00
2	451.72	2024-11-24 00:00:00
3	8525.28	2023-04-06 00:00:00
4	7306.17	2025-01-21 00:00:00

Banking_Loans (2500 rows)

Banking_Cards (4000 rows)

What are the different types of transactions?

The different types of transactions are:

- Deposit
- Withdraw
- Transfer
- Payment

How many transactions happened in total?

The total number of transactions is 20,000.

What is the earliest transaction date in the database?

Type your message...

EYouth
Data Analytics
Bootcamp

CHATBOT

CONCLUSION

EYouth Data Analytics Bootcamp

Project Outcome

We successfully built a full-stack banking analytics system using:

- SQL (for precise data extraction)
- Power BI (for real-time dashboards)
- Python (for advanced analysis & forecasting)
- AI Chatbot (for natural language interaction)

Key Achievements

- Cleaned and modeled data from 6 banking datasets
- Built 5+ dashboards for Customer, Accounts, Loans, and Support
- Applied ML models (Risk classification, Clustering, Anomaly detection)
- Forecasted loan and transaction trends using SARIMAX
- Deployed a chatbot trained on real banking data for live queries

What We Learned

- Data must be prepared carefully to be trusted
- Visuals must answer business questions, not just look pretty
- AI can bridge the gap between raw data and decision-making
- End-to-end integration (SQL → Python → BI → Chatbot) creates real value

Our Great Team

https://linktr.ee/TTeam20

THANK YOU

EYouth Data Analytics Bootcamp

YOUth Learning.com

❯ End Slide