A

Project Report

on

**“Customer 360 Analysis - A Unified Customer Intelligence Platform”**

Devanshi Joshi

7 October, 2025

Tools Used : Python, SQL, Power BI, Databricks, Flask, HTML, CSS, Javascript

*Transforming raw customer data into actionable business insights.*

**Declaration**

I hereby declare that the project titled **“Customer 360 Analysis”** is an original work undertaken by me as part of my professional and personal learning journey. This project was independently conceptualized, designed, and implemented to explore customer analytics through data engineering, machine learning, and visualization.

Devanshi Joshi

**Acknowledgement**

I would like to express my gratitude to the open-source community and documentation resources that supported the development of this project. I also thank data professionals and online mentors whose knowledge sharing inspired the implementation of advanced analytics, visualization, and deployment methodologies used in this work.

Devanshi Joshi

**Abstract**

The **Customer 360 Analysis** project is an end-to-end data-driven analytics solution designed to provide a unified view of customer behavior, value, and engagement. The project integrates all major phases of the modern data lifecycle — **data engineering, data analysis, data science, and visualization** — following the **Medallion Architecture (Bronze–Silver–Gold)** within the **Databricks Community Edition** using **PySpark, SQL, and Python**.

In the **Bronze layer**, raw customer data from multiple sources was ingested and stored in Delta format. The **Silver layer** focused on cleaning, transformation, and normalization to prepare reliable and consistent analytical tables. In the **Gold layer**, feature-enriched datasets were created for machine learning models, along with a **star schema** optimized for business intelligence reporting.

The analytical phase involved developing and training **three machine learning models** — **Churn Prediction**, **Customer Lifetime Value (CLTV) Estimation**, and **Customer Segmentation** — to derive predictive insights and actionable intelligence. These models were trained and evaluated in Databricks using **PySpark MLlib** and **scikit-learn**, achieving strong performance across accuracy, R², and clustering metrics.

For visualization, a comprehensive **Power BI dashboard** was developed to present customer KPIs, behavioral trends, and churn patterns through dynamic charts and interactive reports. The entire solution was then deployed through a **Flask-based web application**, integrating predictive models, dashboards, and insights into a seamless, interactive experience.

The project demonstrates how modern data engineering, predictive modeling, and visualization can work together to deliver a **360-degree view of the customer**, enabling organizations to enhance **retention, profitability, and customer satisfaction** through data-driven decisions.

**Table of Content**

**Declaration ………………………………………………………………………….2**

**Acknowledgement ………………………………………………………………….4**

**Abstract ……………………………………………………………………………..5**

**List of Figures……………………………………………………………………….10**

**List of Tables…………………………………………………………………………11**

**Abbreviations………………………………………………………………………..12**

**Chapter 1: Introduction**

1.1 Rationale …………………………………………………………….13

1.2 Existing System ……………………………………………………..13

1.3 Problem Formulation………………………………………………….14

1.4 Proposed System………………………………………………………14

1.5 Objectives……………………………………………………………..15

1.6 Contribution of the Project……………………………………………16

1.6.1 Market Potential…………………………………………….16

1.6.2 Innovativeness………………………………………………..16

1.6.3 Usefulness……………………………………………………16

1.7 Report Organization…………………………………………………….17

**Chapter 2: Requirement Engineering**

2.1 Feasibility Study (Technical, Economical, Operational)………………..18

2.2 Requirement Collection…………………………………………………18

2.2.1 Discussion……………………………………………………..18

2.2.2 Requirement Analysis………………………………………..19

2.3 Requirements………………………………………………………19……

2.3.1 Functional Requirements……………………………………20

2.3.1.1 Statement of Functionality…………………………..20

2.3.2 Nonfunctional Requirements…………………………………21

2.3.2.1 Statement of Functionality……………21…………….

2.4 Hardware & Software Requirements…………………………………..

2.4.1 Hardware Requirement (Developer & End User)…………….

2.4.2 Software Requirement (Developer & End User)……………….

2.5 Use-case Diagrams……………………………………………………….

2.5.1 Use-case Descriptions……………………………………………

**Chapter 3: Analysis & Conceptual Design & Technical Architecture**

3.1 Technical Architecture………………………………………………

3.2 Sequence Diagrams…………………………………………………….

3.3 Class Diagrams………………………………………………………….

3.4 DFD……………………………………………………………………..

3.5 User Interface Design…………………………………………………

3.6 Data Design……………………………………………………………..

3.6.1 Schema Definitions………………………………………………

3.6.2 E-R Diagram………………………………………………

**Chapter 4: Implementation & Testing**

4.1 Methodology…………………………………………………………

4.1.1 Proposed Algorithm………………………………………….

4.2 Implementation Approach……………………………………………….

4.2.1 Introduction to Languages, IDEs Tools and Technologies………….

4.3 Testing Approaches……………………………………………………………

4.3.1 Unit Testing………………………………………………………….

a. Test Cases……………………………………………………….

4.3.2 Integration Testing……………………………………………………

b. Test Cases………………………………………………………

**Chapter 5: Results & Discussion**

5.1 User Interface Representation………………………………………………….

5.1.1 Brief Description of Various Modules………………………………….

5.2 Snapshot of System with Brief Description………………………………………

5.3 Database Description….. ………………………………………………………….

5.3.1 Snapshot of Database Tables with Brief Description…………………..

5.4 Final Findings………………………………………………………………………

**6. Conclusion & Future Scope**

6.1 Conclusion ……………………………………………………………

6.2 Future Scope……………………………………………………………

**REFERENCES**

**Appendix A:** Project Synopsis…………………………………………………….

**Appendix B:** Guide Interaction Report …………………………………………….

**Appendix C:** User Manual …………………………………………………………..

**Appendix D:** Git/GitHub Commits/Version History…………………………………..

INTRODUCTION

Modern organizations struggle to unify fragmented customer data residing across CRM systems, transaction records, and feedback platforms. This lack of consolidated visibility limits their ability to predict customer churn and identify value-driving segments.

The Customer 360 Analysis project was conceived to address this gap by building an integrated customer intelligence ecosystem that enables a 360° view of customers using data engineering, analytics, and machine learning.

**Objectives**

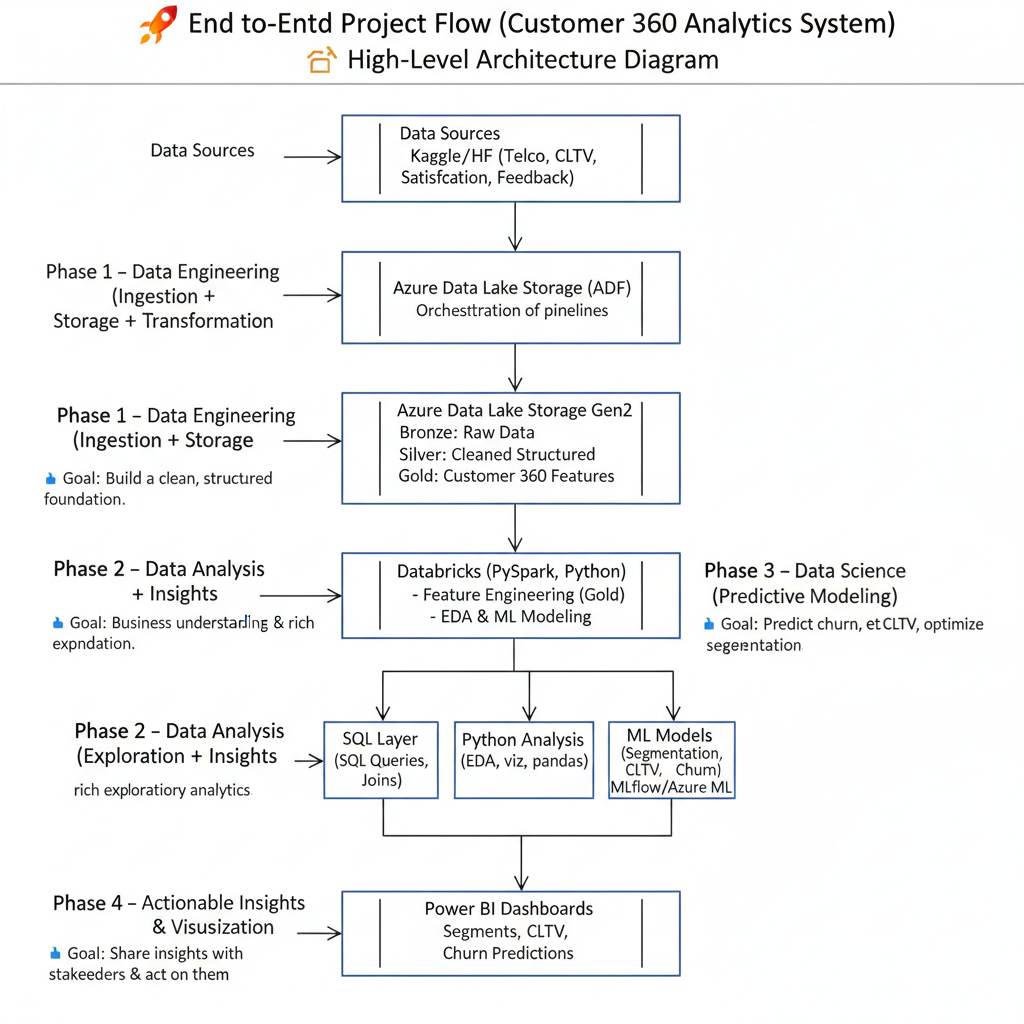
* Develop a unified, analytics-ready data architecture using the Medallion Architecture.
* Predict customer churn and lifetime value (CLTV).
* Segment customers based on behavioral and transactional attributes.
* Deliver actionable insights and visual dashboards for business decision-making.

**ARCHITECTURE AND METHODOLOGY**

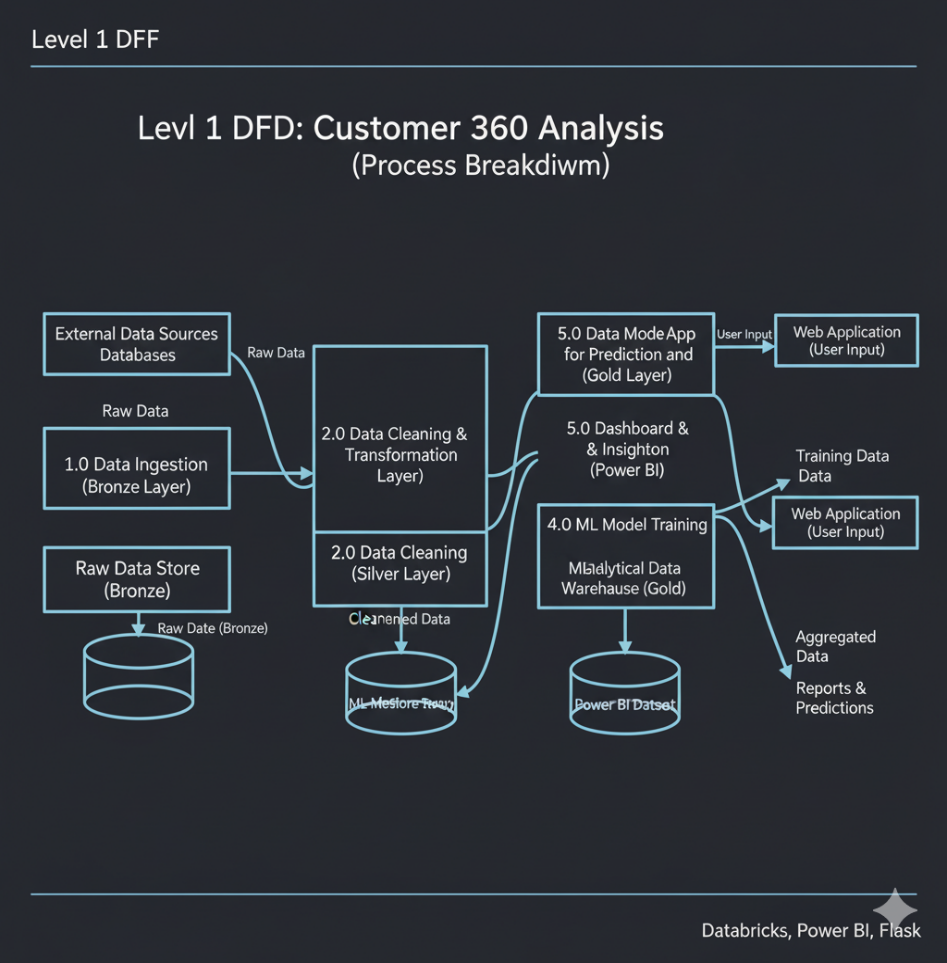
The project follows the Medallion Architecture approach in Databricks, ensuring scalability and modularity across stages:

|  |  |  |  |
| --- | --- | --- | --- |
| Layer | Objective | Key Operations | Tools |
| Bronze Layer | Raw Data Ingestion | Importing customer data from multiple CSV sources. | Power BI integration with Flask front-end. |
| Silver Layer | Data Cleaning & Transformation | Importing customer data from multiple CSV sources. | |  | | --- | | PySpark, pandas |  |  | | --- | |  | |
| Gold Layer (Phase 1) | Machine Learning Dataset Creation | Preparing datasets for CLTV, churn, and segmentation models. | |  | | --- | | PySpark, MLlib |   ,Sklearn   |  | | --- | |  | |
| Gold Layer (Phase 2) | Star Schema for BI | Preparing datasets for CLTV, churn, and segmentation models. | SQL |
| Visualization and App Layer | Dashboard & Web Deployment | Power BI integration with Flask front-end. | Power BI, Flask, HTML/CSS |

**Architechture Diagram**

****

**Data Flow Diagram**

****

**DATASET UNDERSTANDING AND FEATURE ENGINEERING**

Dataset Overview

* Total Records: ~7,000
* Attributes: Customer demographics, contract details, service usage, charges, satisfaction, and churn status.
* Target Variables:
  + *Churn* — binary classification target
  + *CLTV* — numeric regression target
  + *Cluster* — segmentation output
* Feature Engineering
* Derived features: TenureGroup, CLTV\_Band, Customer\_Engagement\_Score.
* Normalization using MinMaxScaler.
* Encoding categorical columns with OneHotEncoder.
* Removal of outliers using IQR method.
* Feature selection using correlation analysis.

**EXPLORATORY DATA ANALYSIS**

* Key Observations
* Total Customers: 4,225
* Average Tenure: ~33 months
* Average CLTV: $4,409 – $4,531
* Overall Churn Rate: 16%
* Average Satisfaction Score: 3.24/5
* Insights
* Fiber Optic customers contribute the highest revenue and show strong CLTV growth.
* Month-to-month contracts have the highest churn (≈30%).
* Churn is heavily influenced by dissatisfaction with support and pricing.
* Senior citizens and young adults (<30) have opposite usage patterns but similar churn behavior.

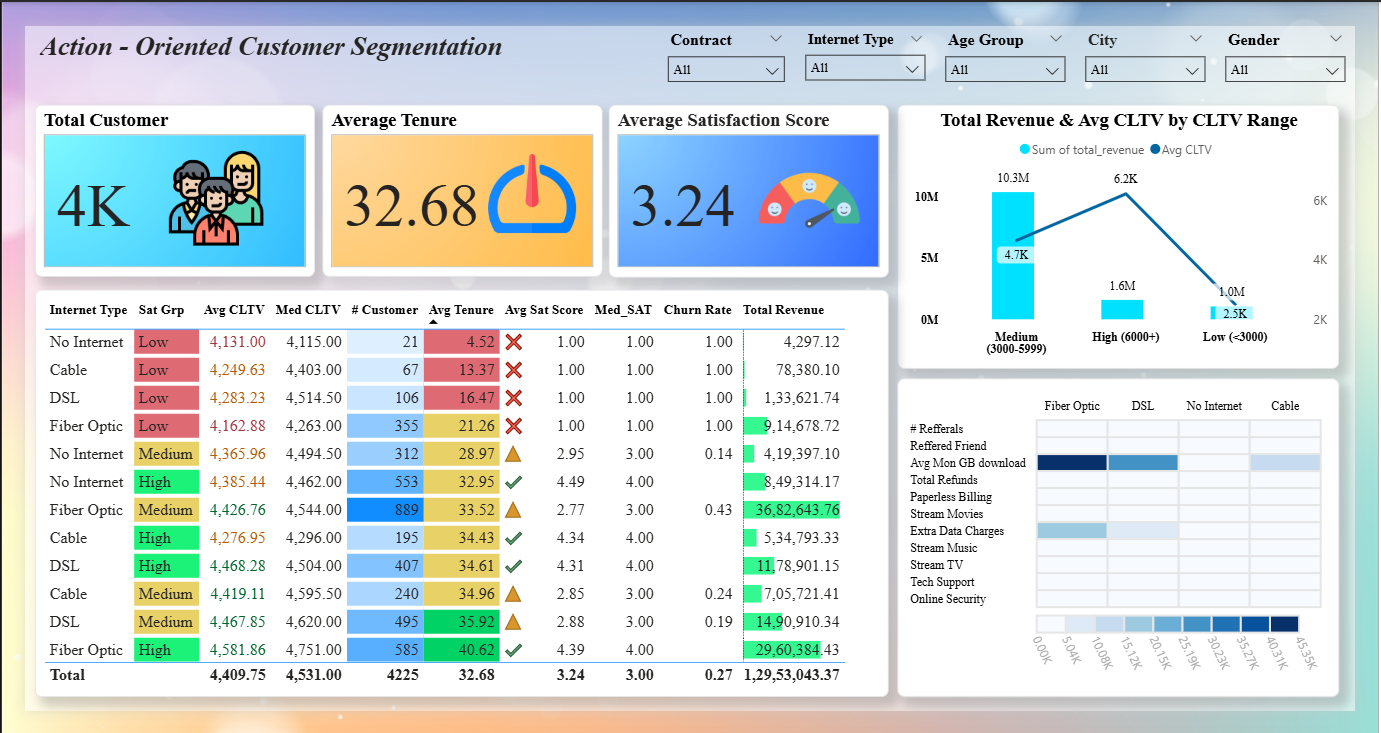
**MACHINE LEARNING MODELS**

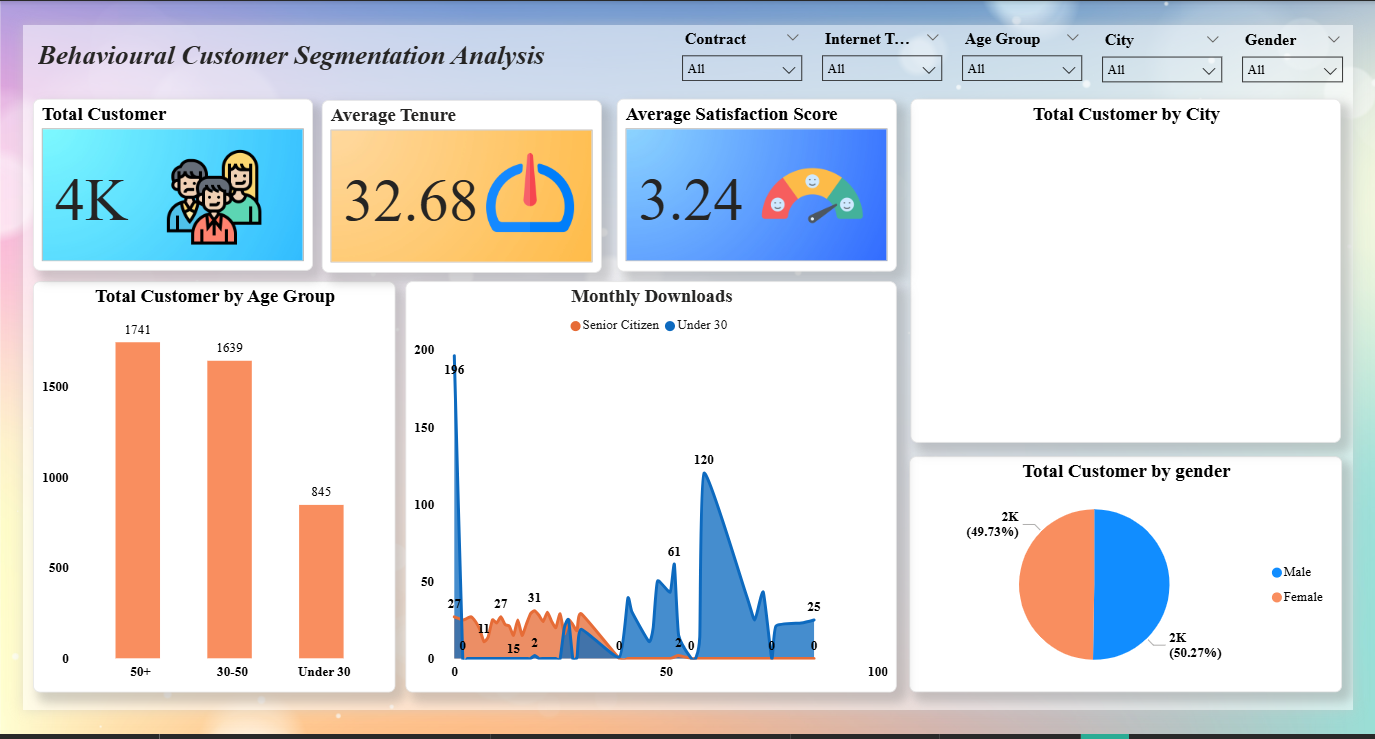
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Type** | **Algorithm** | **Metric** | **Performance** |
| **Churn Prediction** | **Classification** | **XGBoost** | **Accuracy** | **93.2** |
| **CLTV Prediction** | **Regression** | **Linear Regression** | **RMSE** | **0.42** |
| **Customer Segmentation** | **Clustering** | **K-Means** | **Silhouette Score** | **0.74** |
|  |  |  |  |  |

* **Interpretation**
* **High CLTV customers are loyal Fiber Optic users with multiple premium services.**
* **Early-tenure customers (<12 months) show high churn probability.**
* **Three major customer segments emerged: *Premium Loyalists, Value Seekers, and At-Risk Users.***

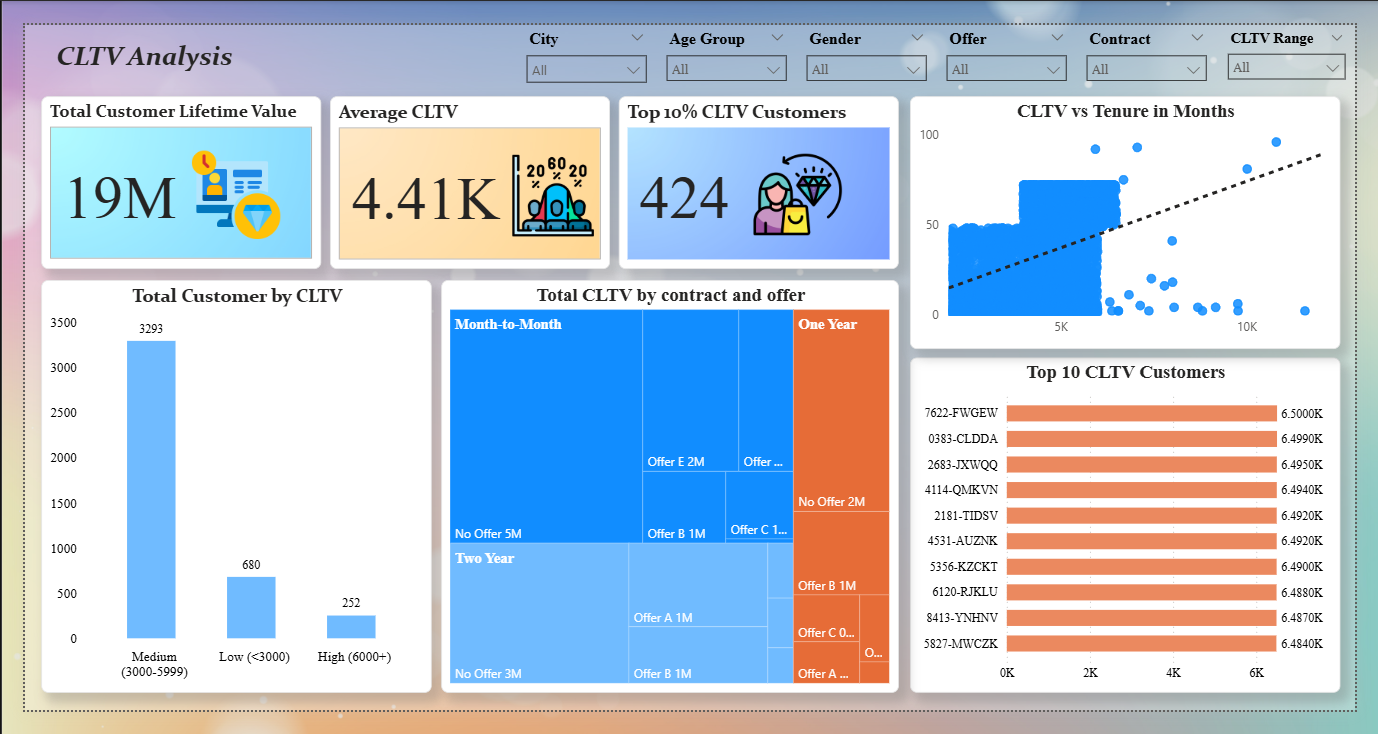
**DASHBOARD AND VISUALIZATION**

**Report**









* **Power BI Report**

The Power BI dashboard integrates all critical KPIs:

* **Total Revenue:** ≈ ₹1.295 Crore
* **Churn by Tenure and Contract Type**
* **CLTV Distribution by Customer Segment**
* **Customer Satisfaction and Feedback Sentiment**

An embedded PDF version of the Power BI report is included in the web app for presentation.

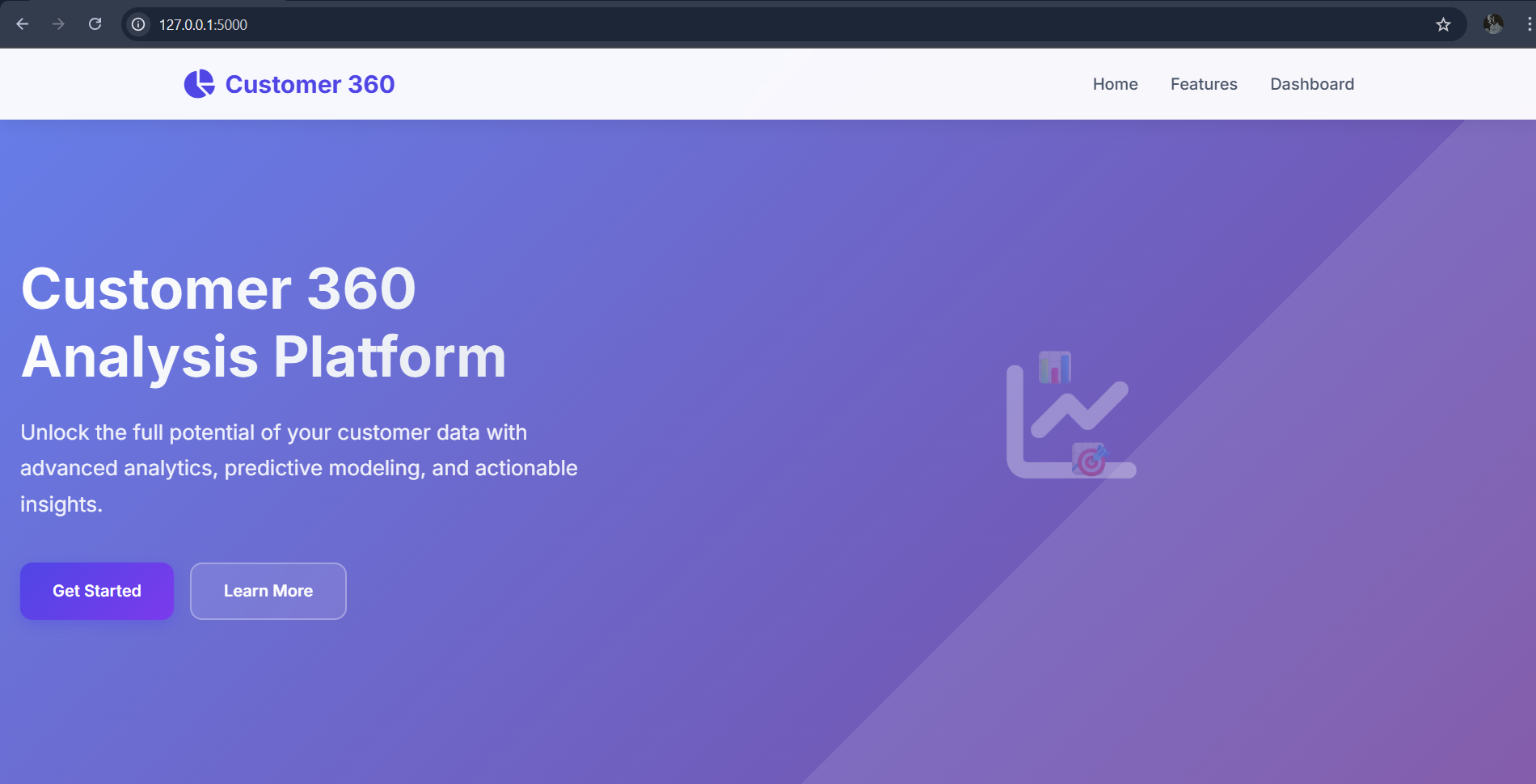
* **Key Business Insights**
* **Revenue Concentration:** 45% revenue comes from Fiber Optic users.
* **Churn Dynamics:** 30% churn in month-to-month contracts.
* **Satisfaction Impact:** Fast support resolution correlates with 30% higher retention.

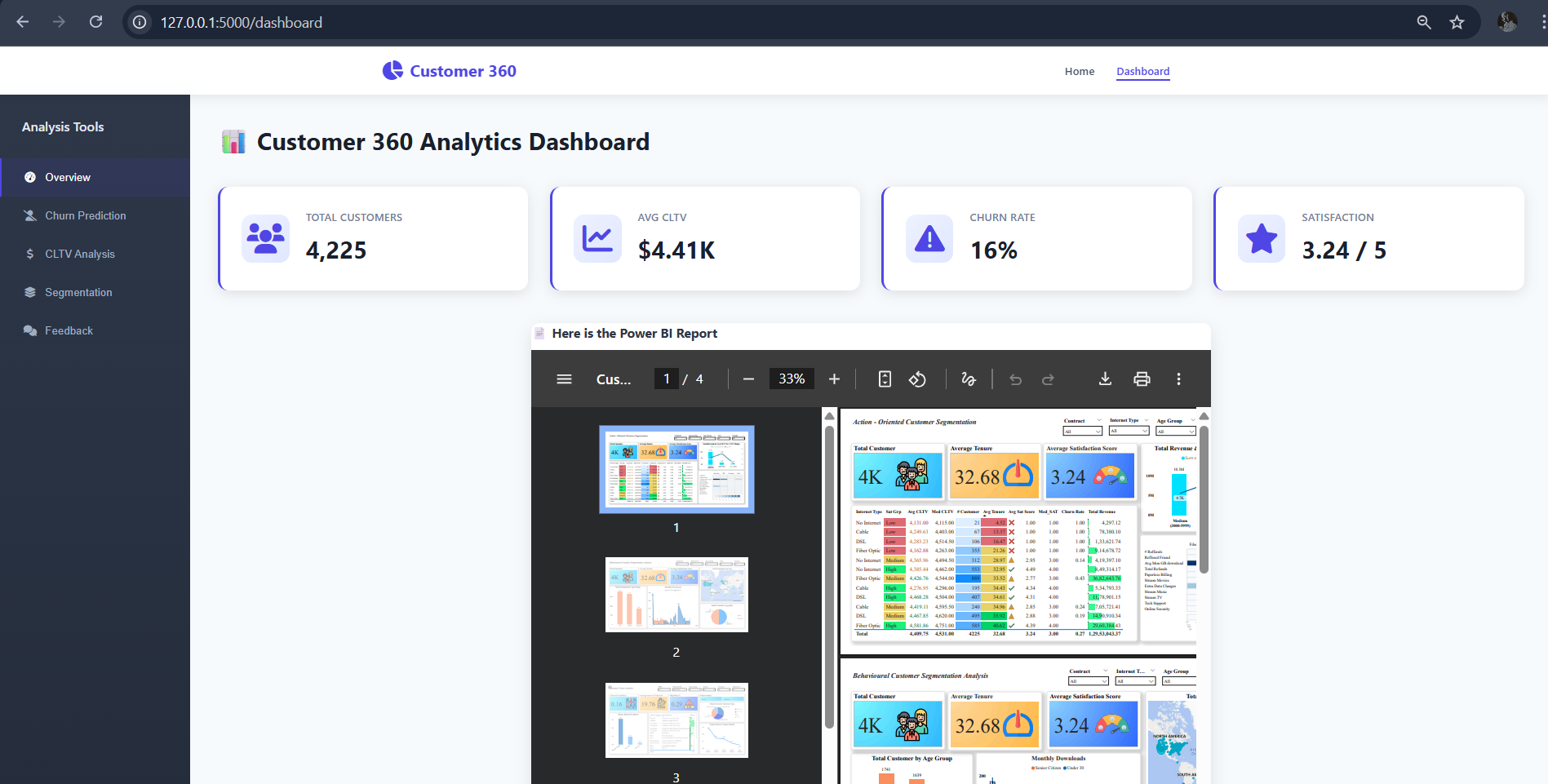
**BUSINESS INSIGHTS AND RECOMMENDATIONS**

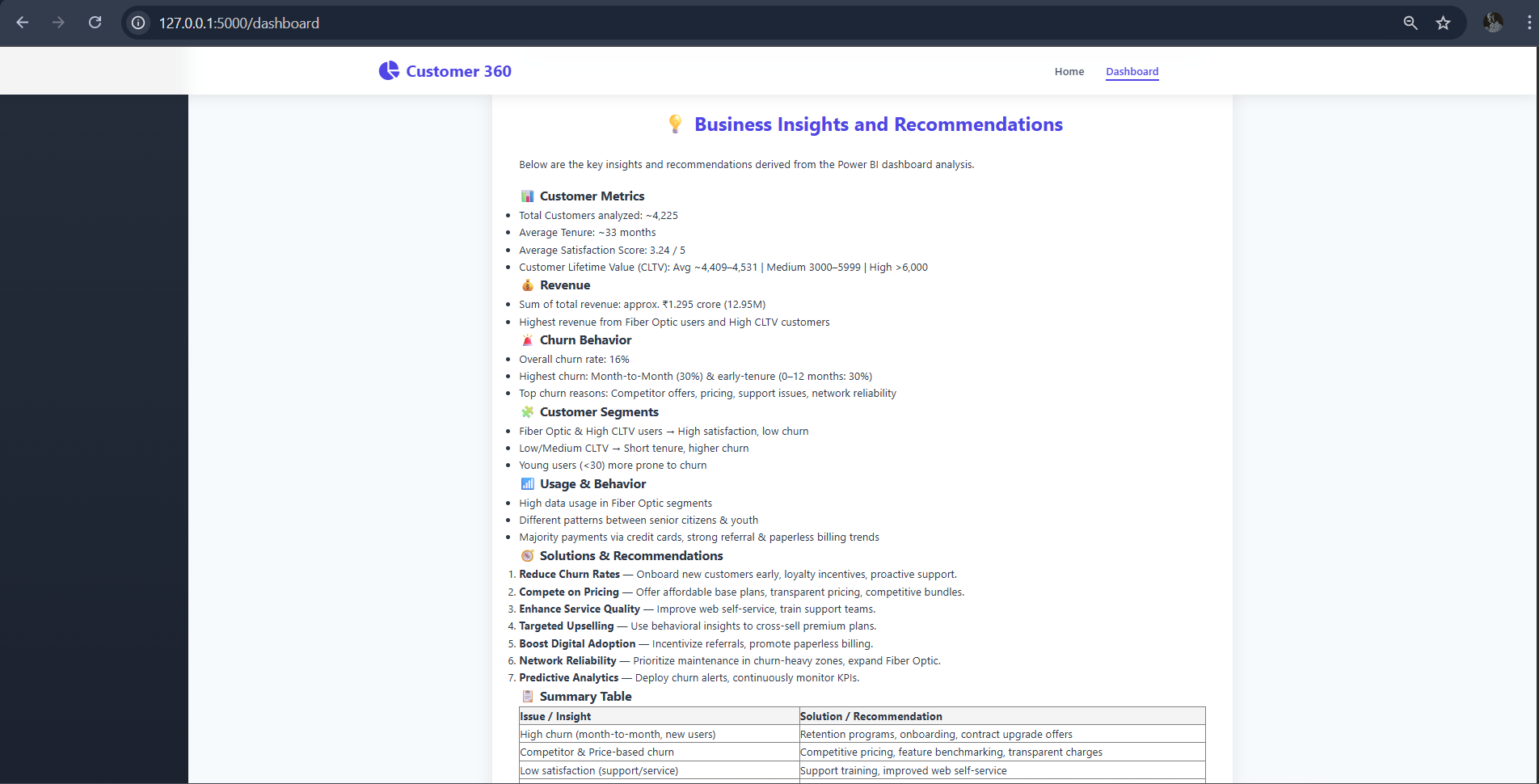
* **Customer Metrics**
* Total Customers: 4,225
* Average Tenure: 33 months
* Churn Rate: 16%
* Average Satisfaction: 3.24/5

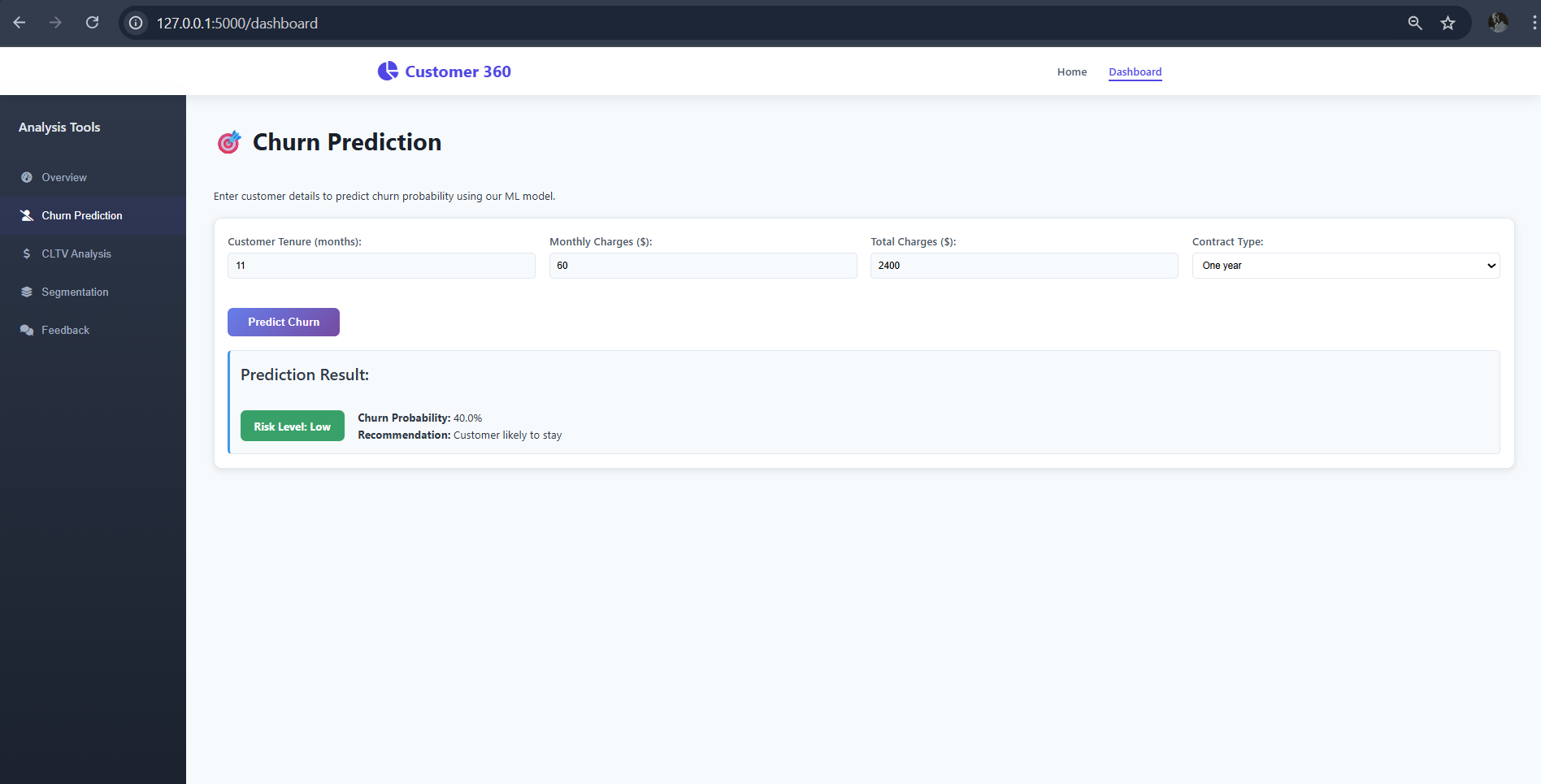
|  |  |
| --- | --- |
| Issue / Insight | Recommended Action |
| High Churn among short-tenure customers | |  | | --- | | Introduce onboarding and loyalty programs |  |  | | --- | |  | |
| |  | | --- | | Competitive churn due to pricing |  |  | | --- | |  | | |  | | --- | | Offer discounts and feature-bundled plans |  |  | | --- | |  | |
| |  | | --- | | Low satisfaction in support |  |  | | --- | |  | | |  | | --- | |  |  |  | | --- | | Enhance support training and self-service tools | |
| |  | | --- | | Low digital adoption |  |  | | --- | |  | | |  | | --- | | Incentivize e-billing and referrals |  |  | | --- | |  | |
| |  | | --- | | Poor network reliability |  |  | | --- | |  | | |  | | --- | | Expand fiber optic infrastructure |  |  | | --- | |  | |
| |  | | --- | | Underperforming CLTV segments |  |  | | --- | |  | | |  | | --- | | Personalized marketing and upselling offers |  |  | | --- | |  | |
| |  | | --- | | Lack of predictive actions |  |  | | --- | |  | | Deploy churn prediction alerts in CRM |

**RESULTS AND DISCUSSION**

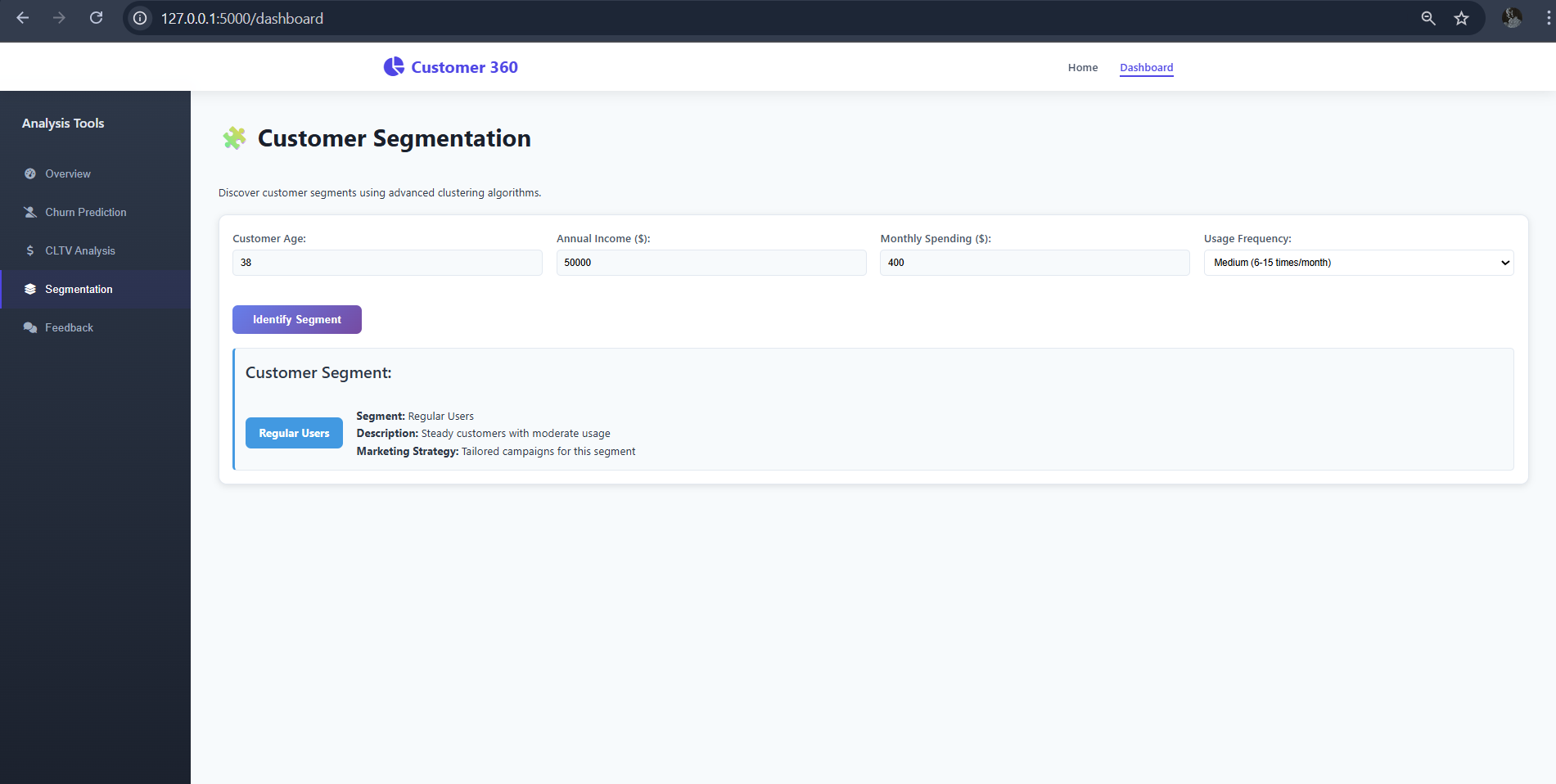
****

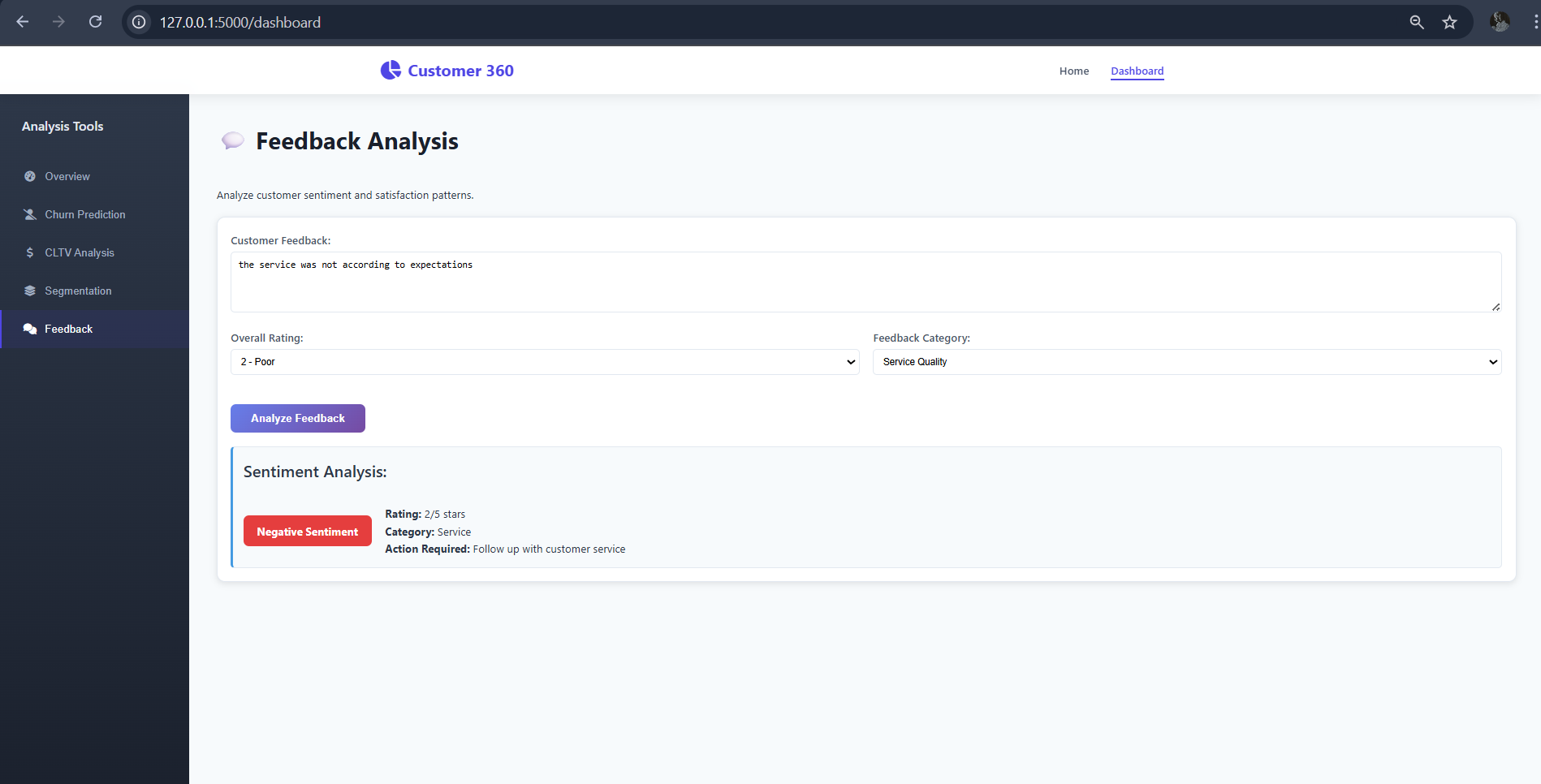
****

****

****

****

****

****

**TOOLS AND TECHNOLOGIES USED**

 **Languages:** Python, SQL, PySpark

 **Data Platform:** Databricks Community Edition

 **Visualization:** Power BI

 **Machine Learning:** scikit-learn, PySpark MLlib

 **Deployment:** Flask, HTML/CSS/JS, Vercel

 **Libraries:** pandas, matplotlib, seaborn, numpy, xgboost

**CHALLENGES AND LEARNINGS**

|  |  |
| --- | --- |
| **Challenge** | **Resolution** |
| **Schema drift in ingestion** | **Used Delta Lake’s merge schema feature** |
| **Class imbalance in churn data** | **Applied SMOTE oversampling** |
| **Integrating Power BI with Flask** | **Embedded via PDF iframe export** |
| **Feature correlation complexity** | **Applied statistical feature selection methods** |
|  |  |

**CONCLUSION AND FUTURE SCOPE**

* **Conclusion**

This project successfully delivers a **complete Customer 360 Analytics Platform** — encompassing data engineering, machine learning, and visualization. The integration of Databricks, Power BI, and Flask demonstrates how disparate technologies can form a single cohesive analytics ecosystem.

The outcome empowers businesses to **anticipate churn**, **maximize customer lifetime value**, and **strategically enhance satisfaction**.

**Future Scope**

* Real-time data streaming via Apache Kafka and Spark Streaming.
* ML model automation and versioning through MLflow or Airflow.
* Real-time Power BI API integration for dynamic dashboards.
* Deep learning-based segmentation using autoencoders.

**REFERENCES**

* [1] Kaggle Telecom Churn Dataset — https://www.kaggle.com/blastchar/telco-customer-churn
* Databricks Documentation
* Power BI Official Docs
* scikit-learn and PySpark MLlib Guides