

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, R^2 , 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.

1. 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.

1.1 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.

1.2 Problem Statement

In many organizations, customer data is stored in multiple disconnected systems such as CRM, billing, feedback, and service databases. This fragmentation limits the ability to perform comprehensive analysis and obtain a complete view of customer behavior.

As a result, businesses face challenges such as:

- Inability to accurately predict customer churn and identify risk segments.
- Difficulty in estimating customer lifetime value (CLTV) for revenue forecasting.
- Lack of data-driven segmentation for personalized marketing and retention efforts.
- Limited actionable insights due to absence of integrated dashboards and analytics.

This project aims to solve these challenges by creating an **end-to-end Customer 360 Analysis platform** that integrates data from multiple sources, performs predictive analytics, and visualizes key business insights.

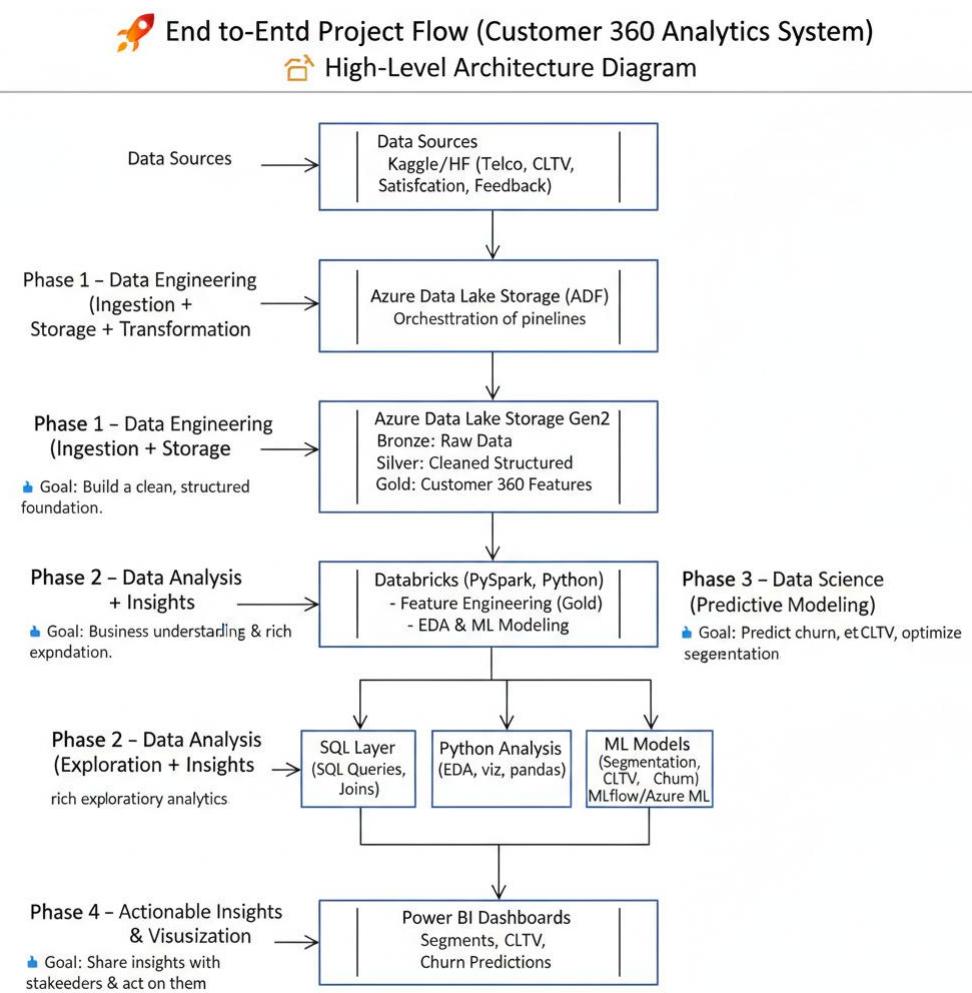
2. ARCHITECTURE AND METHODOLOGY

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

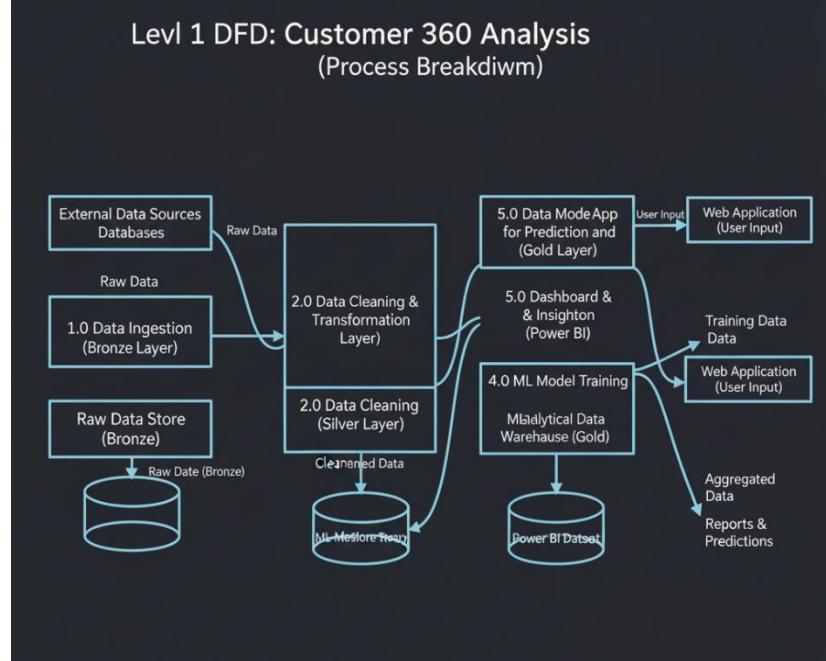
Layer	Objective	Key Operations	Tools
Bronze Layer	Raw Data Ingestion	Importing customer data from multiple CSV sources.	Power BI integration with Flask front-end.

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
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

2.1 Architecture Diagram



2.2 Data Flow Diagram



3. DATASET UNDERSTANDING AND FEATURE ENGINEERING

3.1 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

3.2 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.

4. EXPLORATORY DATA ANALYSIS

4.1 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

4.2 Insights

- Fiber Optic customers contribute the highest revenue and show strong CLTV growth.
- Month-to-month contracts have the highest churn ($\approx 30\%$).
- Churn is heavily influenced by dissatisfaction with support and pricing.
- Senior citizens and young adults (<30) have opposite usage patterns but similar churn behavior.

5. MACHINE LEARNING MODELS

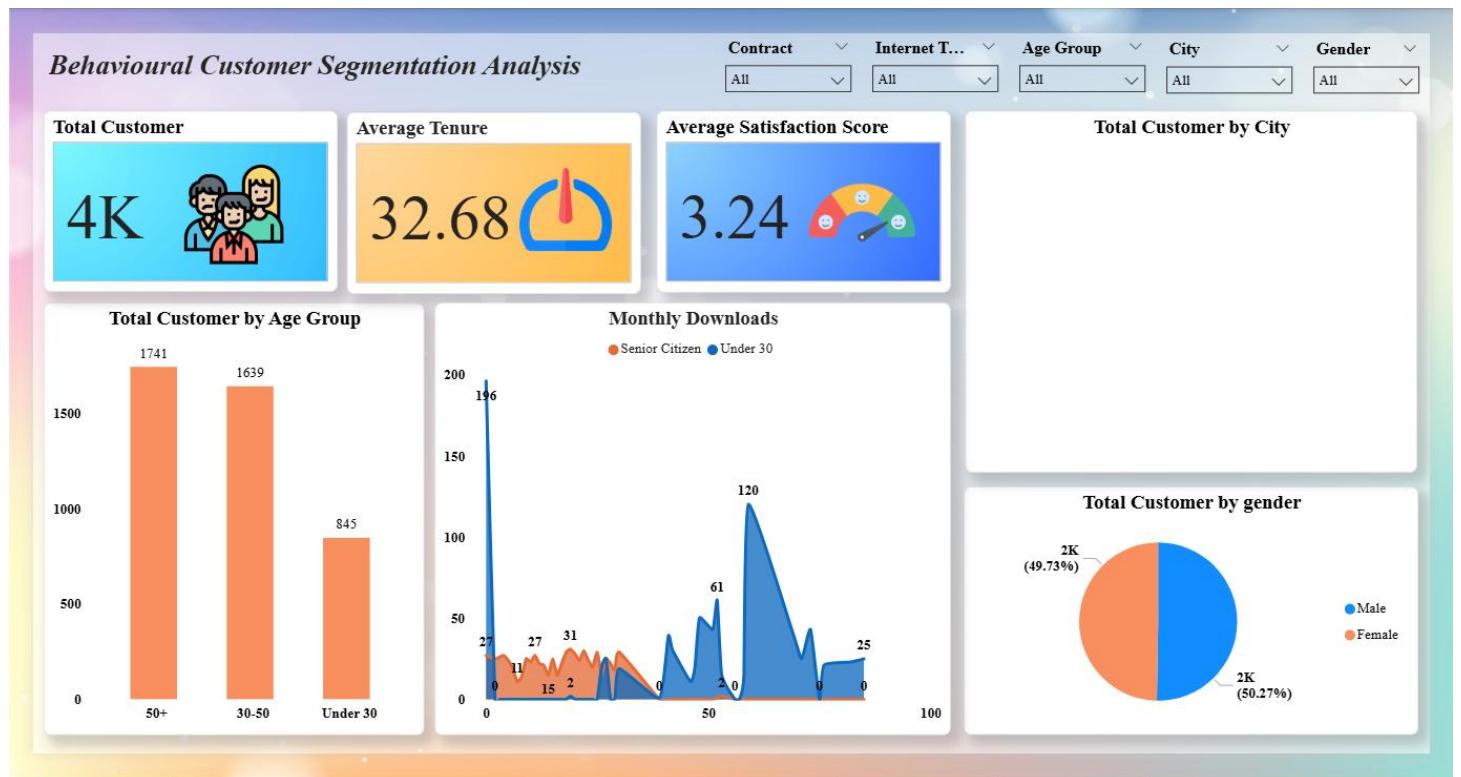
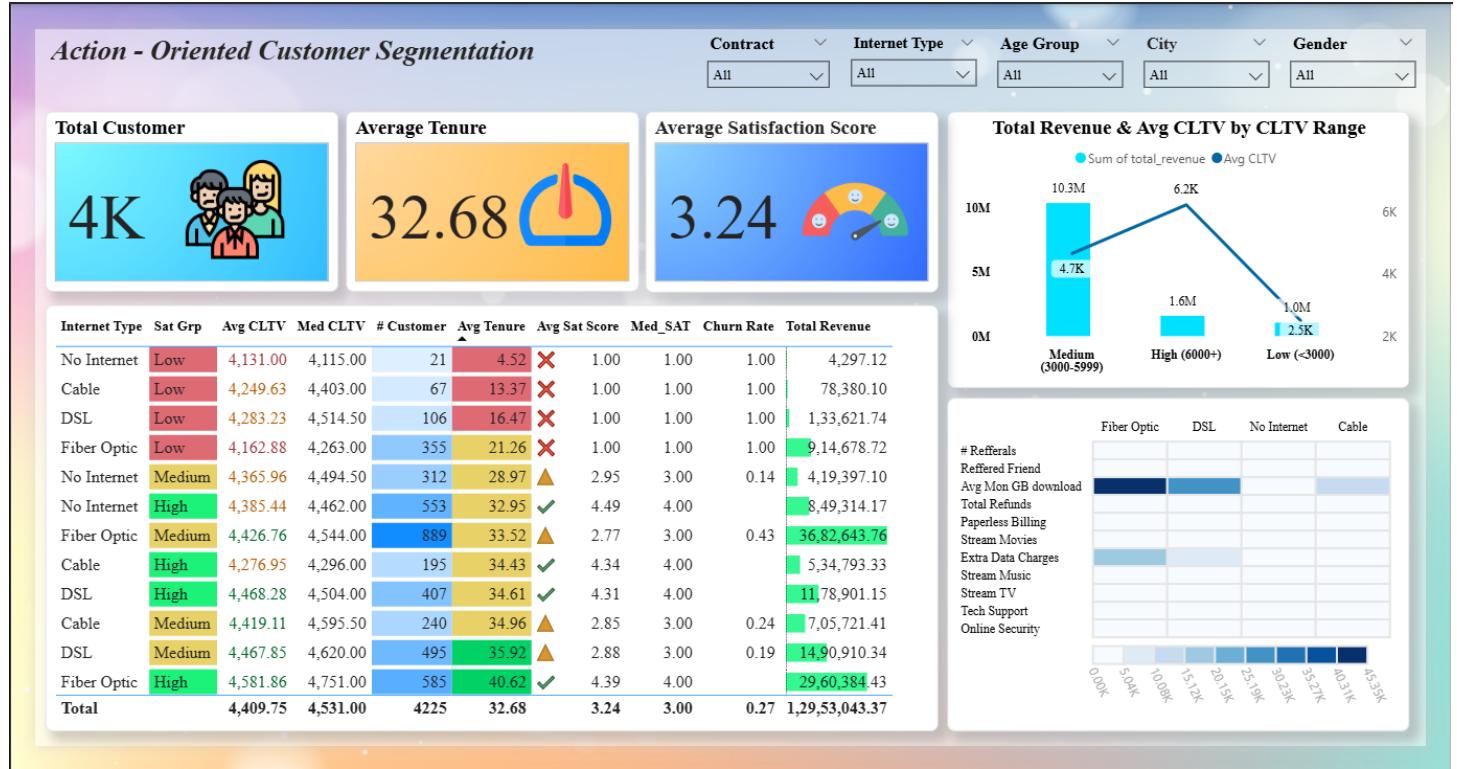
Model	Type	Algorithm	Metric	Performance
Churn Prediction	Classification	XGBoost	Accuracy	93.2
CLTV Prediction	Regression	Linear Regression	RMSE	0.42
Customer Segmentation	Clustering	K-Means	Silhouette Score	0.74

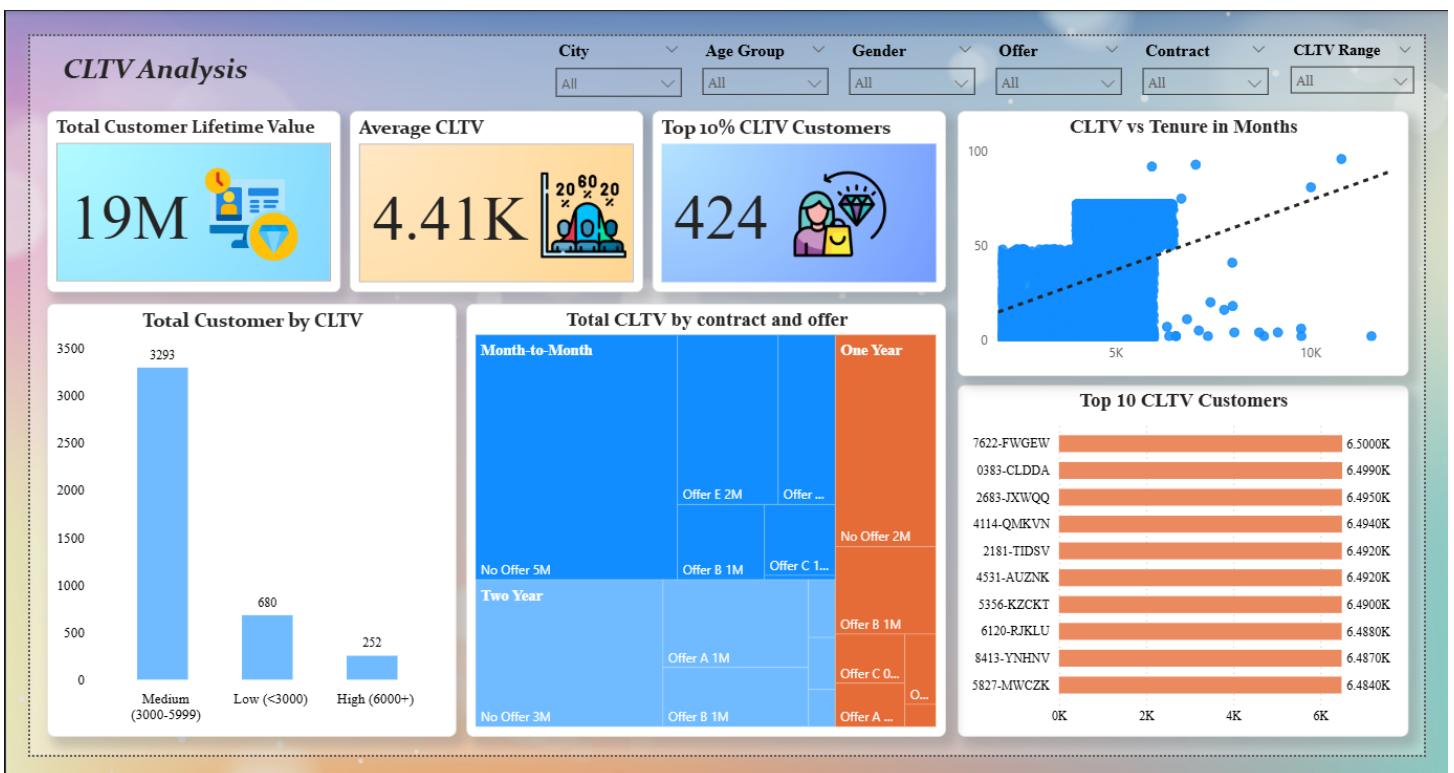
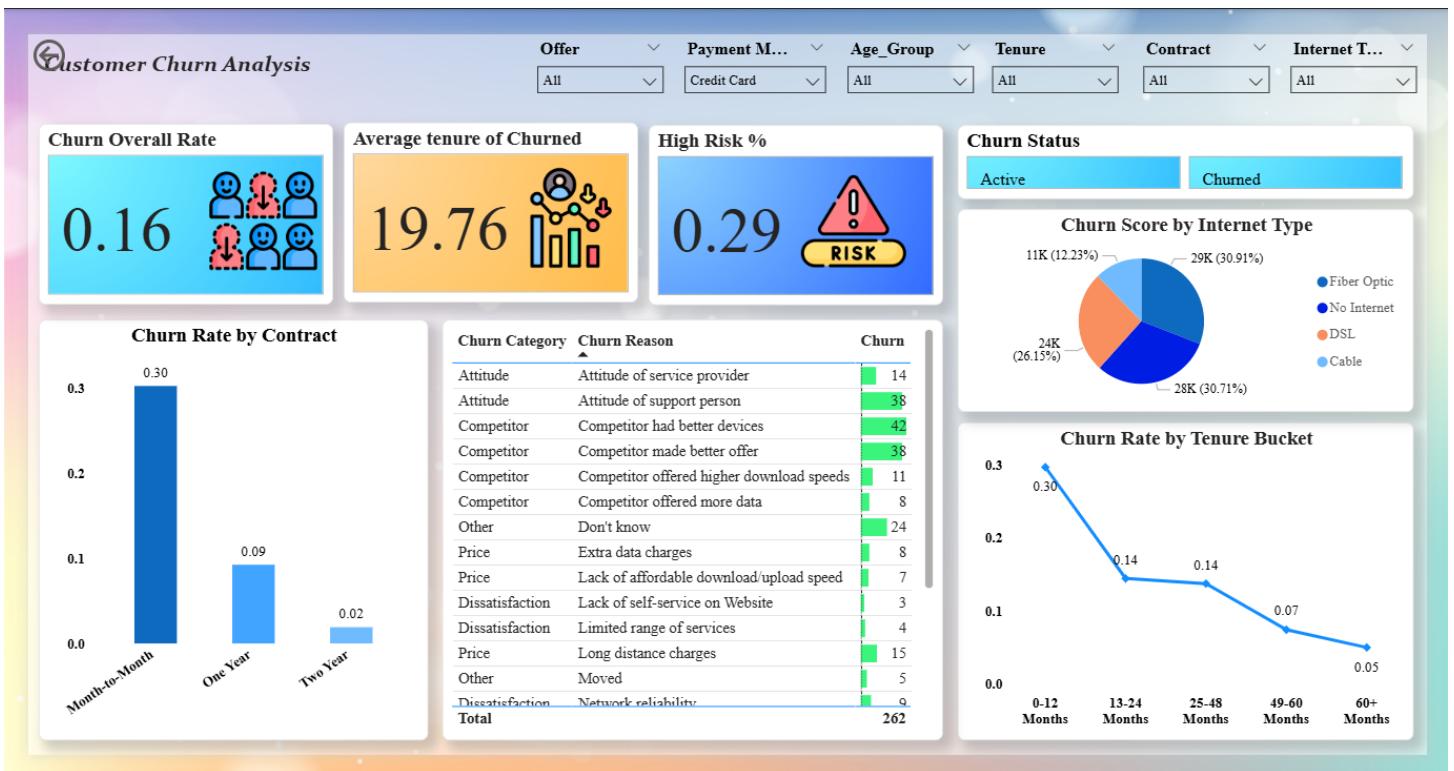
5.1 Interpretation

- 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*.

6. DASHBOARD AND VISUALIZATION

6.1 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.

6.2 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.

7. BUSINESS INSIGHTS AND RECOMMENDATIONS

1. Customer Metrics & Revenue Overview

- Total Customers Analyzed: ~4,225
- Average Tenure: ~33 months
- Average Satisfaction Score: 3.24 / 5
- Average CLTV (Customer Lifetime Value): ₹4,409–₹4,531 (medium range ₹3,000–₹5,999; high-value > ₹6,000)
- Total Revenue: ₹12.95M (\approx ₹1.295 crore)

Insights:

- Majority of revenue (~45%) is driven by high CLTV customers, primarily Fiber Optic users.
- Customers with longer tenure and multiple premium services show significantly higher CLTV and lower churn probability.

Recommendation:

- Focus retention programs and loyalty rewards on Fiber Optic and long-tenure customers.
- Develop premium upsell bundles for medium CLTV customers to move them into the high-value segment.

2. Churn Analysis

- Overall Churn Rate: 16%
- High-Risk Segments: Month-to-Month contract customers (~30% churn)
- Tenure 0–12 months: 3x higher churn risk than others.
- Common Reasons for Churn: Pricing dissatisfaction, poor service support, and inconsistent network quality.

Insights:

- Early-tenure customers are most vulnerable to churn.
- Low satisfaction (<3) and short contracts correlate strongly with churn.
- Customers who contacted support multiple times in a short period had a 25% higher churn rate.

Recommendations:

- Introduce welcome engagement programs and first-year loyalty discounts.

- Provide personalized retention offers for customers flagged as at-risk by churn prediction model.
- Monitor service quality KPIs and trigger intervention when dissatisfaction or complaints rise.

3. Customer Segmentation

Three major behavioral clusters identified:

1. Premium Loyalists — Long tenure, high CLTV, multiple premium services, low churn.
2. Value Seekers — Mid CLTV, price-sensitive, moderate satisfaction, high upsell potential.
3. At-Risk Users — Short tenure, low satisfaction, month-to-month contracts, high churn risk.

Insights:

- Value Seekers show openness to add-on services like streaming bundles if discounted.
- At-Risk Users show low adoption of digital and referral programs.

Recommendations:

- Use personalized campaigns to upsell streaming and device protection plans to Value Seekers.
- Offer contract upgrade incentives to At-Risk Users.
- Maintain exclusive perks for Premium Loyalists to sustain engagement.

4. Satisfaction & Feedback Trends

- Average Feedback Score: 3.2/5
- Customers with negative sentiment have a 2× higher churn rate.
- Faster issue resolution correlates with 30% higher retention.

Insights:

- Service quality and response time directly impact retention.
- Sentiment analysis reveals support and billing issues as primary dissatisfaction causes.

Recommendations:

- Deploy AI-powered sentiment monitoring on customer feedback data.
- Set up automated escalation workflows for low-satisfaction responses.
- Regularly train customer service teams on soft skills and issue resolution.

5. Usage & Behavior

- Average Monthly Data Usage (GB): Significantly higher for Fiber Optic users.
- Senior citizens use less data but have longer tenure and higher loyalty.
- Younger users (<30) show higher churn and low contract renewals.

Recommendations:

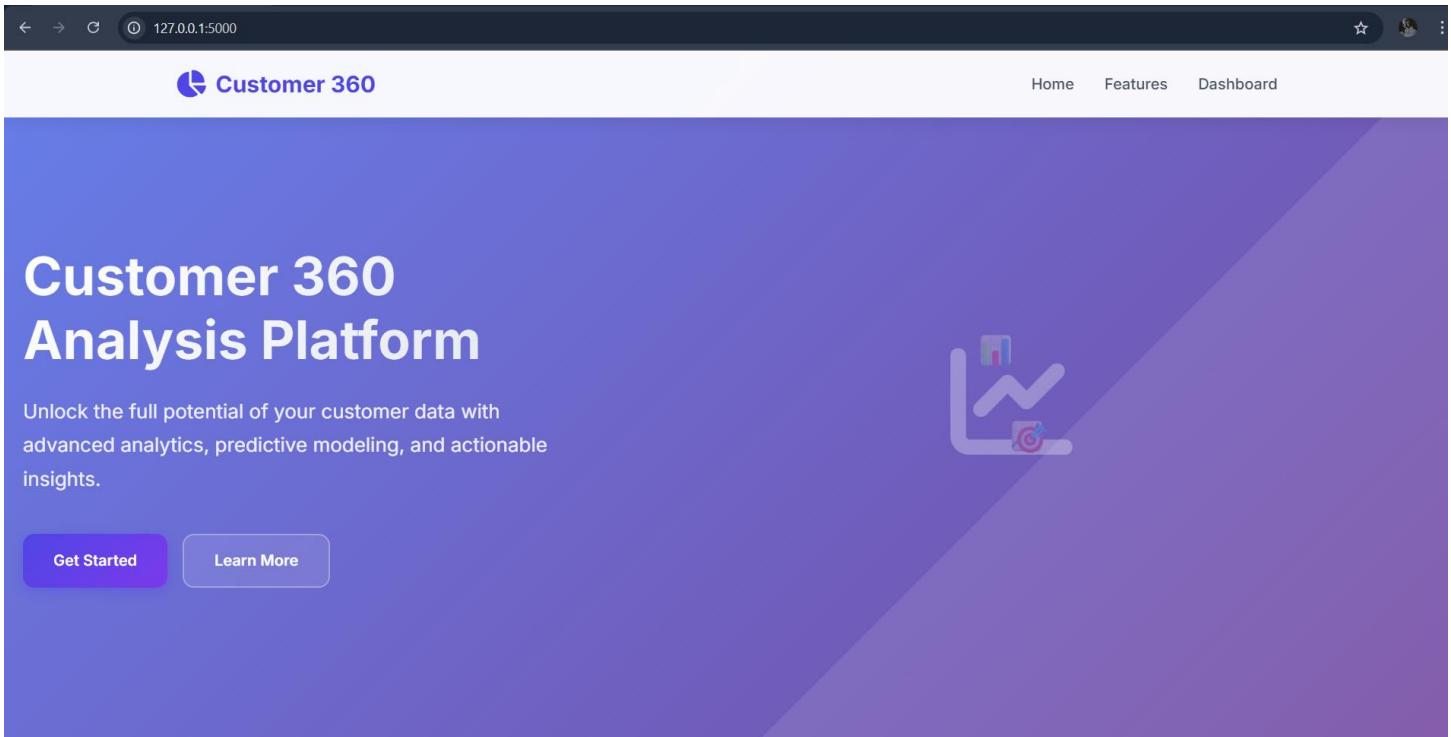
- Personalize offers based on usage (e.g., data add-ons for heavy users).
- Create “youth loyalty” programs with flexible pricing and digital perks.

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

8. RESULTS AND DISCUSSION

8.1 UI design



The screenshot shows the homepage of the Customer 360 Analysis Platform. At the top, there is a dark header bar with browser controls (back, forward, search) and a URL field showing "127.0.0.1:5000". The main title "Customer 360" is displayed with a blue icon. Below the title, the platform's name "Customer 360 Analysis Platform" is written in large white text. A subtext below it reads: "Unlock the full potential of your customer data with advanced analytics, predictive modeling, and actionable insights." Two buttons are visible at the bottom left: a purple "Get Started" button and a white "Learn More" button. On the right side, there is a graphic of a chart with bars and a target symbol.

Customer 360 Analytics Dashboard

TOTAL CUSTOMERS 4,225

AVG CLTV \$4.41K

CHURN RATE 16%

SATISFACTION 3.24 / 5

Here is the Power BI Report

Action-Oriented Customer Segmentation

Segment	Total Customers	Avg Tenure	Avg Satisfaction Score
Total Customer	4K	32.68	3.24
Cable	4,225	44.03	3.24
Fiber Optic	4,225	43.14	3.24
No Internet	4,225	44.94	3.24
Other	4,225	43.01	3.24
Young	4,225	44.70	3.24
Older	4,225	45.30	3.24
High Income	4,225	44.94	3.24
Low Income	4,225	44.70	3.24
Male	4,225	44.94	3.24
Female	4,225	44.70	3.24
Contract	4,225	44.94	3.24
Internet	4,225	44.70	3.24
Age Group	4,225	44.82	3.24
Total Revenue	4,225	44.82	3.24

Behavioral Customer Segmentation Analysis

Segment	Total Customers	Avg Tenure	Avg Satisfaction Score
Total Customer	4K	32.68	3.24
Cable	4,225	44.03	3.24
Fiber Optic	4,225	43.14	3.24
No Internet	4,225	44.94	3.24
Other	4,225	43.01	3.24
Young	4,225	44.70	3.24
Older	4,225	45.30	3.24
High Income	4,225	44.94	3.24
Low Income	4,225	44.70	3.24
Male	4,225	44.94	3.24
Female	4,225	44.70	3.24
Contract	4,225	44.94	3.24
Internet	4,225	44.70	3.24
Age Group	4,225	44.82	3.24
Total Revenue	4,225	44.82	3.24

Business Insights and Recommendations

Below are the key insights and recommendations derived from the Power BI dashboard analysis.

Customer Metrics

- Total Customers analyzed: ~4,225
- Average Tenure: ~33 months
- Average Satisfaction Score: 3.24 / 5
- Customer Lifetime Value (CLTV): Avg ~4,409–4,531 | Medium 3000–5999 | High >6,000

Revenue

- Sum of total revenue: approx. ₹1.295 crore (12.95M)
- Highest revenue from Fiber Optic users and High CLTV customers

Churn Behavior

- Overall churn rate: 16%
- Highest churn: Month-to-Month (30%) & early-tenure (0–12 months: 30%)
- Top churn reasons: Competitor offers, pricing, support issues, network reliability

Customer Segments

- Fiber Optic & High CLTV users → High satisfaction, low churn
- Low/Medium CLTV → Short tenure, higher churn
- Young users (<30) more prone to churn

Usage & Behavior

- High data usage in Fiber Optic segments
- Different patterns between senior citizens & youth
- Majority payments via credit cards; strong referral & paperless billing trends

Solutions & Recommendations

- Reduce Churn Rates — Onboard new customers early, loyalty incentives, proactive support.
- Compete on Pricing — Offer affordable base plans, transparent pricing, competitive bundles.
- Enhance Service Quality — Improve web self-service, train support teams.
- Targeted Upselling — Use behavioral insights to cross-sell premium plans.
- Boost Digital Adoption — Incentivize referrals, promote paperless billing.
- Network Reliability — Prioritize maintenance in churn-heavy zones, expand Fiber Optic.
- Predictive Analytics — Deploy churn alerts, continuously monitor KPIs.

Summary Table

Issue / Insight	Solution / Recommendation
High churn (month-to-month, new users)	Retention programs, onboarding, contract upgrade offers
Competitor & Price-based churn	Competitive pricing, feature benchmarking, transparent charges
Low satisfaction (support/service)	Support training, improved web self-service

Customer 360

Home Dashboard

Churn Prediction

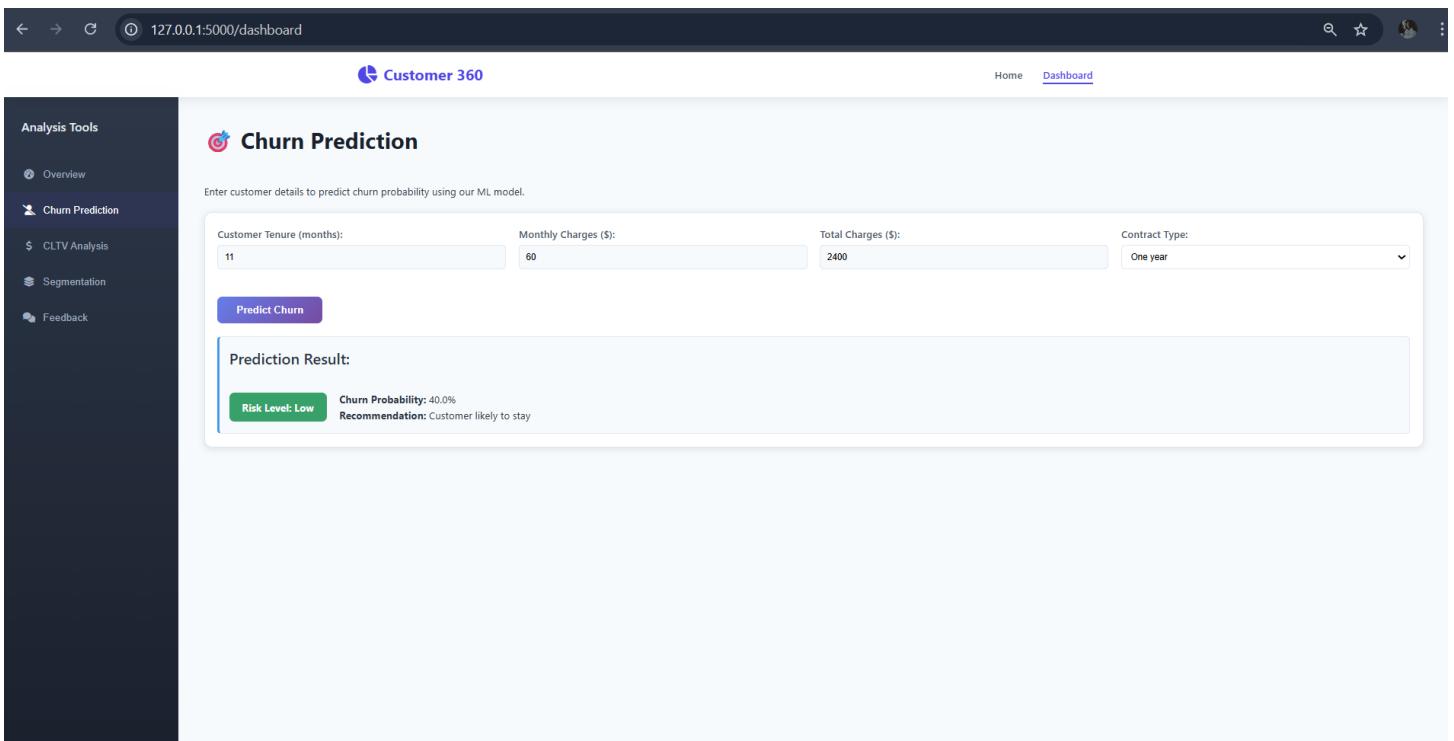
Enter customer details to predict churn probability using our ML model.

Customer Tenure (months): 11 Monthly Charges (\$): 60 Total Charges (\$): 2400 Contract Type: One year

Predict Churn

Prediction Result:

Risk Level: Low Churn Probability: 40.0% Recommendation: Customer likely to stay



Customer 360

Home Dashboard

CLTV Analysis

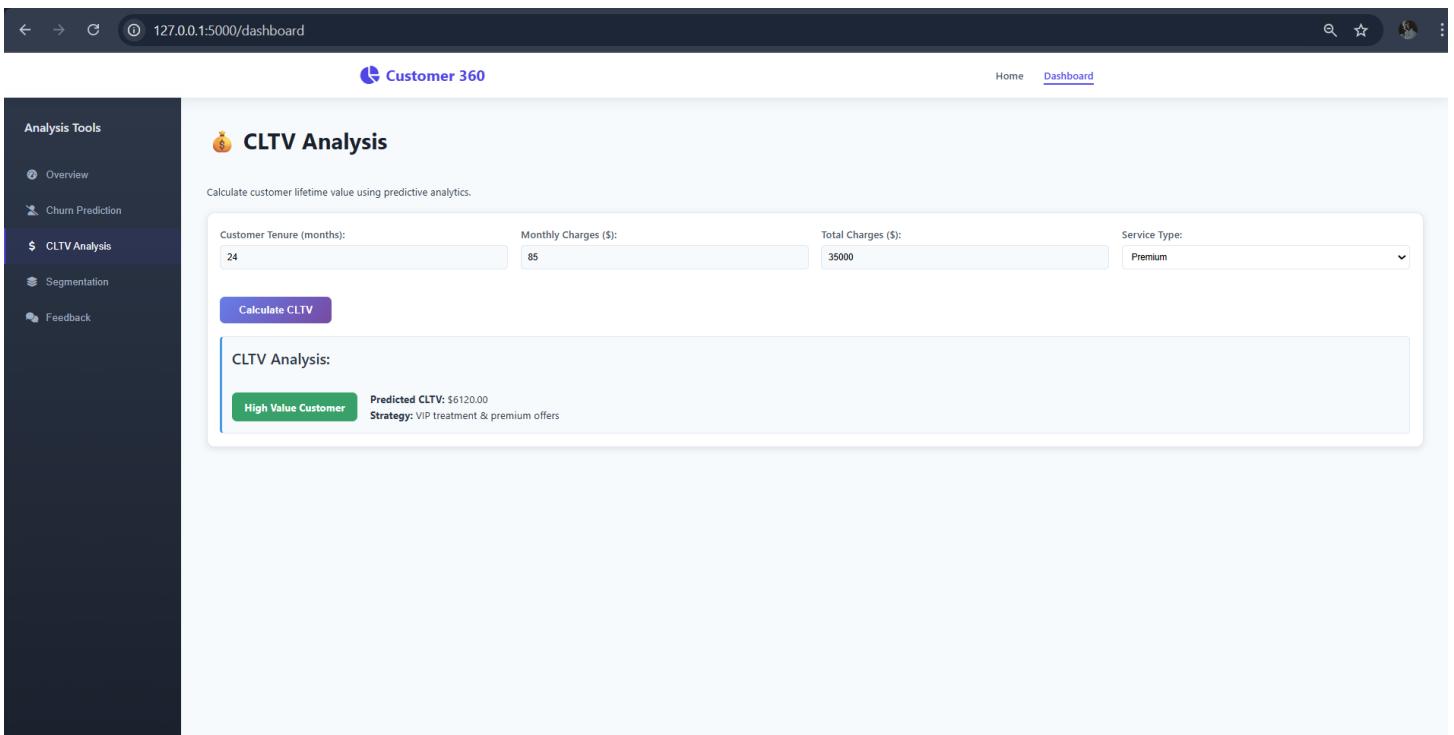
Calculate customer lifetime value using predictive analytics.

Customer Tenure (months): 24 Monthly Charges (\$): 85 Total Charges (\$): 35000 Service Type: Premium

Calculate CLTV

CLTV Analysis:

High Value Customer Predicted CLTV: \$6120.00 Strategy: VIP treatment & premium offers



The screenshot shows the Customer 360 dashboard with the 'Segmentation' tool selected. The main area displays 'Customer Segmentation' with a sub-section titled 'Discover customer segments using advanced clustering algorithms.' It includes input fields for 'Customer Age' (38), 'Annual Income (\$)' (50000), 'Monthly Spending (\$)' (400), and 'Usage Frequency' (Medium (6-15 times/month)). A purple button labeled 'Identify Segment' is present. Below it, a section titled 'Customer Segment:' shows a result for 'Regular Users' with the following details: Segment: Regular Users, Description: Steady customers with moderate usage, and Marketing Strategy: Tailored campaigns for this segment.

The screenshot shows the Customer 360 dashboard with the 'Feedback' tool selected. The main area displays 'Feedback Analysis' with a sub-section titled 'Analyze customer sentiment and satisfaction patterns.' It includes a text input field for 'Customer Feedback' containing the text 'the service was not according to expectations'. Below it, there are dropdown menus for 'Overall Rating' (set to '2 - Poor') and 'Feedback Category' (set to 'Service Quality'). A purple button labeled 'Analyze Feedback' is present. Below it, a section titled 'Sentiment Analysis:' shows a result for 'Negative Sentiment' with the following details: Rating: 2/5 stars, Category: Service, and Action Required: Follow up with customer service.

9. TOOLS AND TECHNOLOGIES USED

- A. **Languages:** Python, SQL, PySpark
- B. **Data Platform:** Databricks Community Edition
- C. **Visualization:** Power BI

- D. **Machine Learning:** scikit-learn, PySpark MLlib
- E. **Deployment:** Flask, HTML/CSS/JS, Vercel
- F. **Libraries:** pandas, matplotlib, seaborn, numpy, xgboost

10. 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

11. CONCLUSION AND FUTURE SCOPE

11.1 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.

11.2 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

For Detailed view of project please follow <https://github.com/Ddevanshi10/Customer360Analysis>

Mail : <mailto:joshidevanshi1012@gmail.com>

LinkedIn : <https://www.linkedin.com/in/devanshi10/recent-activity/all/>