

Space for Attempting Question Paper

Answer 1 :

Background

Customer churn is a significant challenge for businesses in the telecommunications industry. It directly impacts revenue and profitability, making churn prediction an essential part of customer retention strategies. The ability to predict churn using machine learning can help businesses proactively engage with customers and reduce attrition rates.

Introduction

App - *Telco ChurnXpert: A Smart Churn Prediction & Analytics Tool*

The **Telco ChurnXpert** application is a **smart churn prediction and analytics tool** that helps telecom companies identify customer churn patterns and improve customer retention strategies. The application allows businesses to explore customer demographics, visualize service usage trends, assess churn risk, and interactively filter data to derive meaningful insights.

Telco ChurnXpert addresses the critical challenge of customer attrition in the telecom industry. With businesses losing millions annually due to preventable churn, this Streamlit-powered application bridges the gap between predictive analytics and actionable retention strategies.

Below is the Link for Source code and Streamlit App

GitHub- [SAkriti03/TelcoChurnXpert: Smart Prediction tool for Churn](#)

Streamlit App Link - [Telco ChurnXpert · Streamlit](#)

Objectives The key objectives are:

- Develop a predictive model that accurately forecasts customer churn.
- Identify key factors influencing customer retention.
- Provide an interactive dashboard for real-time churn analysis.
- Enable businesses to formulate proactive customer engagement strategies.

Data Processing and Preprocessing

Dataset Overview

The dataset used in this project consists of **7,043 customer records** and contains features that describe customer demographics, service usage, and account-related details.

The dependent variable is **Churn**, which we aim to predict based on the independent variables.

The dataset has:

- **Categorical features** (e.g., gender, Contract, PaymentMethod)
- **Numerical features** (e.g., tenure, MonthlyCharges, TotalCharges)
- **Binary features** (e.g., Churn, Partner, Dependents, PhoneService)

The key attributes include:

- **Customer Demographics:** Gender, Senior Citizen status, Partner, Dependents.
- **Account Information:** Tenure (in months), Monthly Charges, Total Charges, Payment Method, Contract Type.
- **Service Usage:** Phone Service, Internet Service, Online Security, Streaming Services.
- **Churn Indicator:** Whether the customer left the service or remain

Data Cleaning and Transformation

Before using the dataset, preprocessing steps were performed to clean and encode the data.

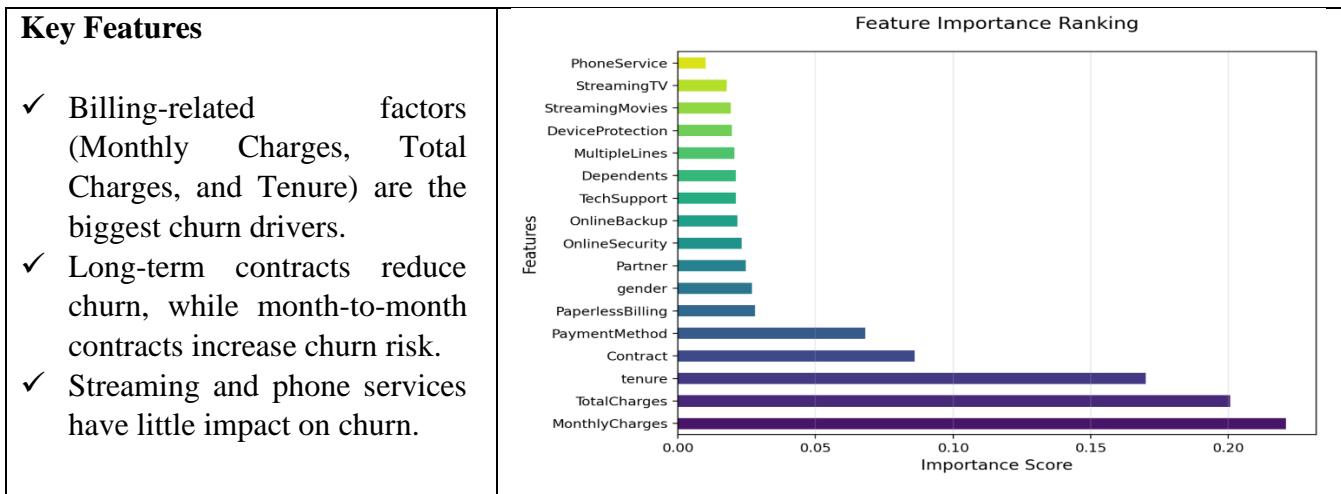
The following transformations were applied:

1. **Removing irrelevant columns:** customerID was dropped as it does not contribute to churn prediction.
2. **Handling missing values:** TotalCharges had missing values, which were replaced with the **median** value.
3. **Converting categorical variables:**
 - Binary features (gender, Partner, Dependents, PhoneService, and PaperlessBilling) were encoded as 0 (No) and 1 (Yes).
 - Multi-category features like Contract were mapped to numeric values (e.g., *Month-to-month* → 0, *One year* → 1, *Two year* → 2).
 - Payment methods were encoded (Electronic check → 0, Mailed check → 1, Bank transfer → 2, Credit card → 3).
4. **Handling "No internet service" values:** These were standardized to "No", ensuring consistency.
5. **Convert the target variable (Churn) into numeric format,** representing "Yes" as 1 and "No" as 0.
6. **Feature scaling:** The Total Charges column was converted to numeric values to avoid format inconsistencies.
7. **Return the cleaned and processed dataset,** ready for machine learning models.

Feature selection

The feature importance scores are derived from the Random Forest Classifier. The **Random Forest Classifier** assigns importance scores to features based on their contribution to model accuracy.

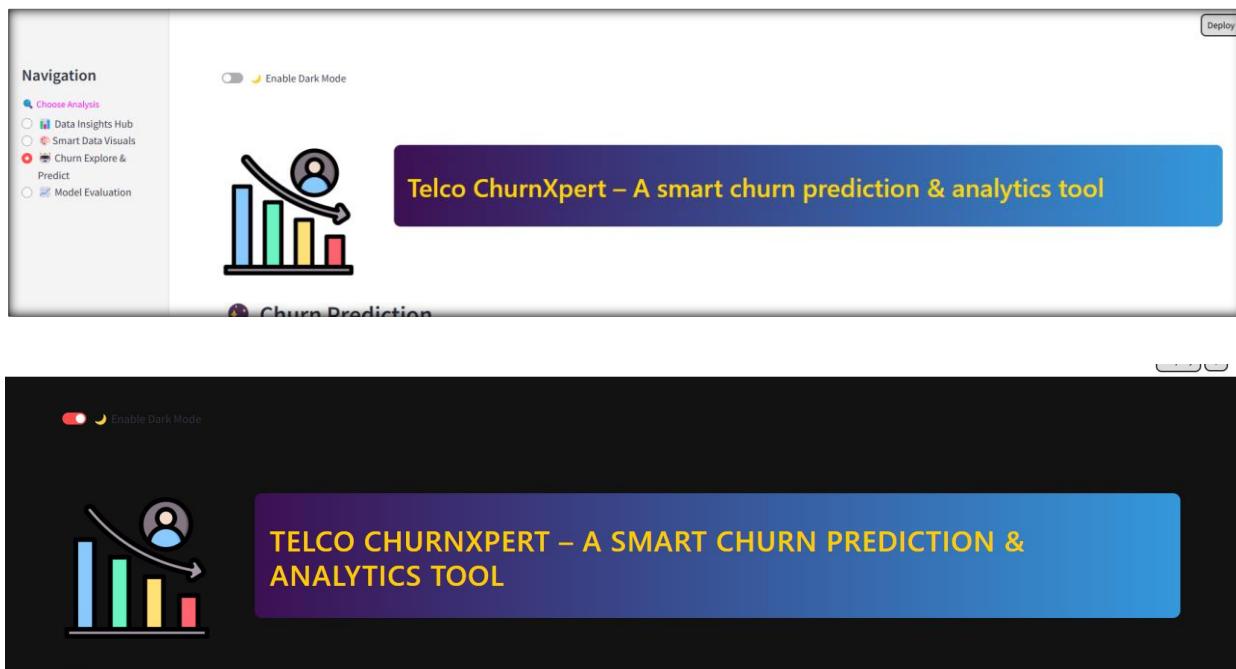
The below graph visualizes the impact of different features on the customer churn prediction



Streamlit Application Features

The application is designed with a **clean UI** and a **sidebar for navigation**. The user can choose between different sections:

A **toggle button** was included to switch between **Dark Mode** and **Light Mode** dynamically. So now App is available in both Mode.



Navigation and Functionalities -The application contains four main sections:

Section	Features
Data Insights Hub	Overview of business metrics and customer data
Smart Data Visuals	Interactive visualizations of churn factors
Churn Explore & Predict	Real-time churn prediction based on customer inputs
Client Persona Evaluation	Key factors influencing retention and attrition.
Model Evaluation	Model performance metrics and feature analysis

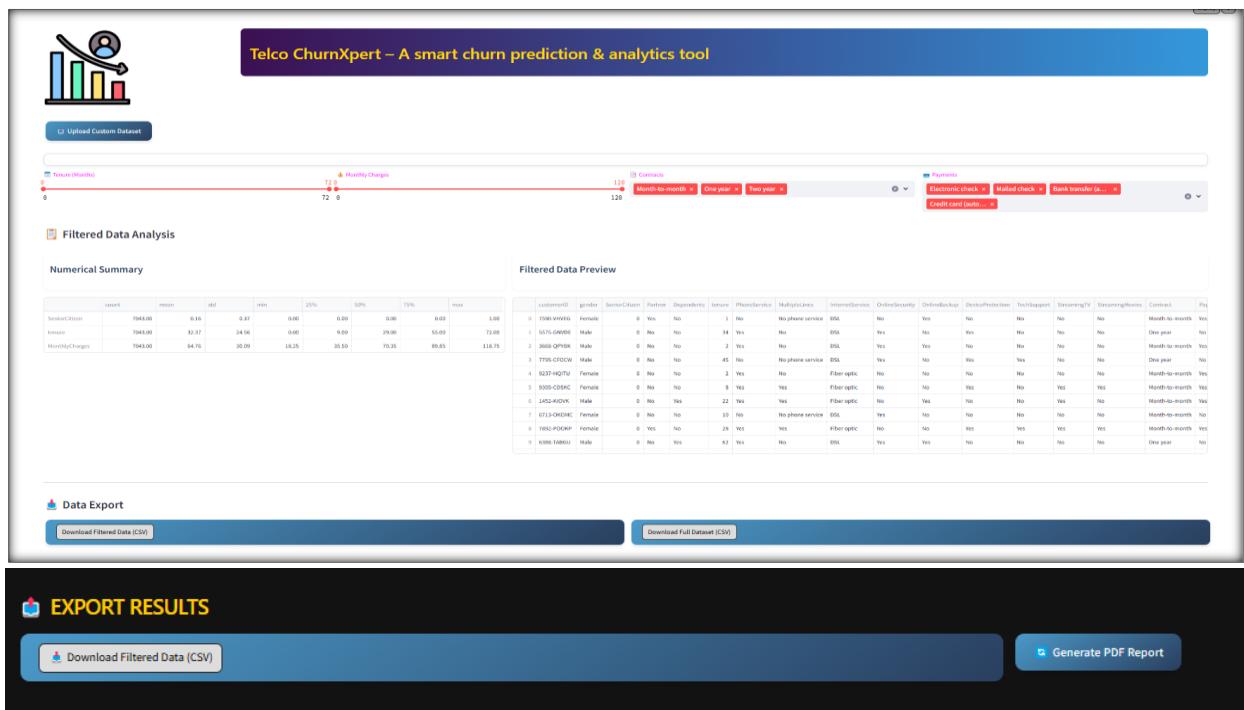
Page1 :Data Insights Hub

The **Telco ChurnXpert** interface provides an intuitive platform for analysing customer churn in the telecommunications sector. The dashboard enables users to filter data, analyse key parameters, and export insights for further decision-making and customer churn prediction in the telecom industry, allowing businesses to:

- Identify high-risk customers and implement retention strategies.
- Understand the impact of contract terms and payment methods on customer loyalty.
- Perform deep-dive analytics by exporting and processing data in external tools.

Customer Analytics Dashboard of the Telco ChurnXpert application provides an interactive interface for exploring customer data, analyzing key metrics, and applying filters to segment customers.

The below dashboard provides valuable insights into customer behavior, tenure, and revenue generation, helping businesses make data-driven retention strategies to reduce churn and increase profitability.



Below is the detailed breakdown of the insights derived from the dashboard.

- **Upload Custom Dataset:** This feature allows users to input their own telecom customer data, ensuring flexibility in analysis.
- **Numerical Summary:** - Displays key statistics such as average tenure, monthly charges, and senior citizen count to understand customer demographics and behavior.

Customer Analytics Dashboard

Customer Analytics Dashboard of the Telco ChurnXpert application provides an interactive interface for exploring customer data, analyzing key metrics, and applying filters to segment customers. Below is a detailed explanation of its business, technical, and functional aspects.

- **Filter Controls** – Allows users to filter data based on tenure, monthly charges, contract type, and payment method.

- Real-time Metrics – Displays key business insights such as total customers, average tenure, churn rate, and average customer lifetime value (CLV).
- Data Explorer – An interactive table displaying customer records.
- Data Filtering Options : The dashboard provides interactive filtering options to refine datasets based on critical variables affecting customer churn:

Filter	Description
Interactive Data Filtering	Users can filter data based on tenure, monthly charges, contract type, and payment methods to refine analysis.
Tenure & Monthly Charges Filter	Helps segment customers based on their subscription duration and billing amount.
Contract Type Filter	Allows analysis of churn trends based on contract duration (Month-to-Month, One-Year, Two-Year).
Payment Method Filter	Identifies churn patterns based on different payment methods like Electronic Check, Mailed Check, Bank Transfer, and Credit Card.

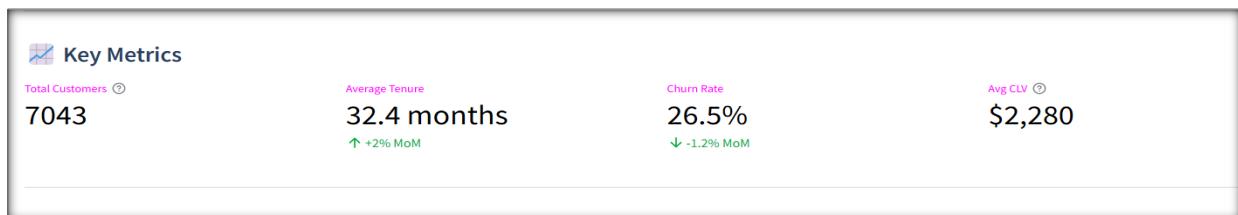
- **Data Export Functionality** At the bottom of the dashboard, two **CSV download options** allow users to extract filtered data or the full dataset for further external analysis

Page2 :Smart Data Visuals

This page of the Telco ChurnXpert tool provides a high-level real-time summary of key customer churn metrics along with an interactive chart explorer for deeper analysis.

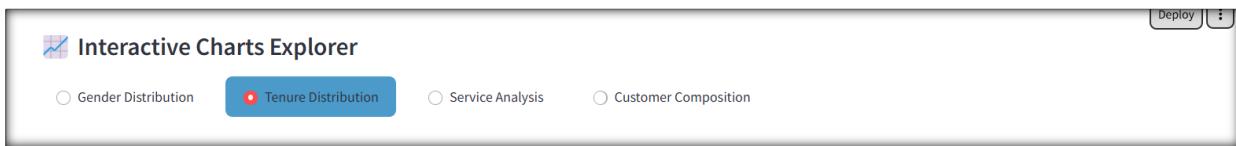
Key Business Metrics: Understanding Customer Churn Trends

The dashboard highlights four critical business indicators that directly impact churn rates and revenue:



Metric	Value	Insight
Total Customers	7,043	The dataset contains a diverse sample of customers, providing a broad analysis.
Average Tenure	32.4 months	Customers stay, on average, for about 2.7 years. The upward trend (+2% MoM) suggests improving retention efforts.
Churn Rate	26.5%	Over one-fourth of customer's churn. The slight decline (-1.2% MoM) indicates that retention strategies are showing some success.
Average CLV (Customer Lifetime Value)	\$2,280	This value reflects the total revenue a business can expect per customer. A higher CLV is crucial for profitability.

Interactive Charts explores :-

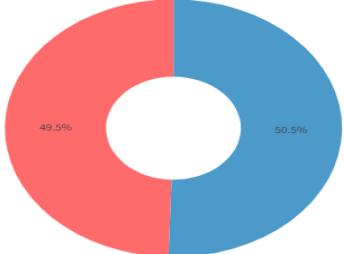
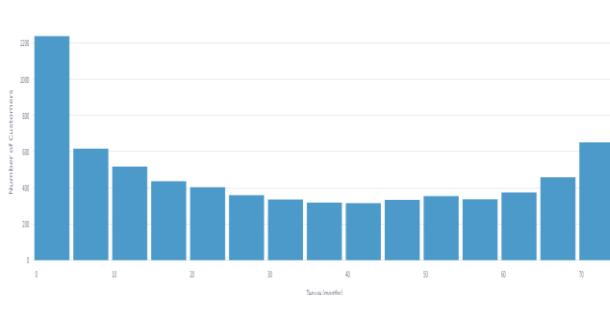
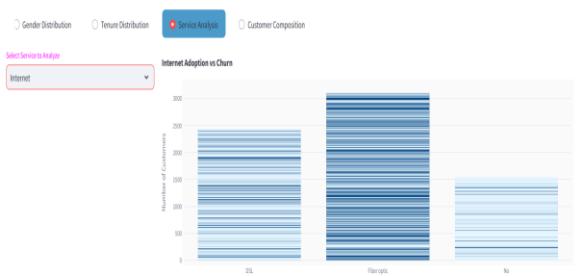


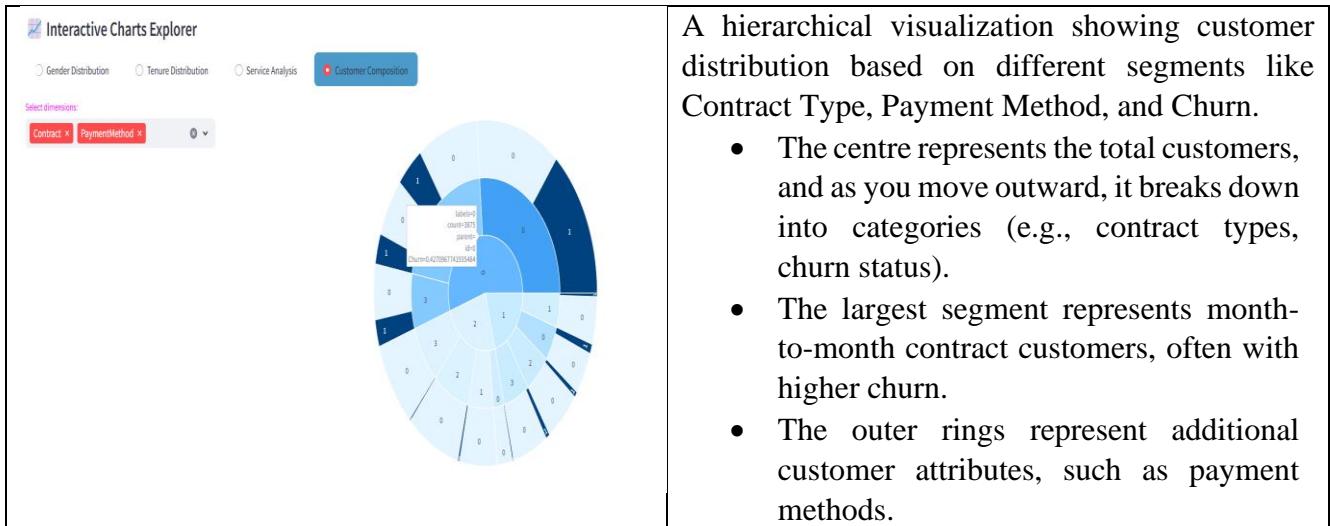
The screenshot shows the 'Interactive Charts Explorer' interface. At the top, there is a title bar with the text 'Interactive Charts Explorer' and a 'Deploy' button. Below the title bar, there are four tabs: 'Gender Distribution' (grey), 'Tenure Distribution' (blue, indicating it is selected), 'Service Analysis' (grey), and 'Customer Composition' (grey). The main content area is currently empty, suggesting the chart for the selected tab is not yet loaded.

Interactive Charts Explorer section allows users to analyse customer data visually. The available chart options include:

- Gender Distribution – Shows the proportion of male and female customers using a donut chart.
- Tenure Distribution – Helps visualize customer retention trends over time.
- Service Analysis – Identifies how different telecom services impact customer churn.
- Customer Composition – Breaks down customer segments based on key demographic and behavioural attributes.

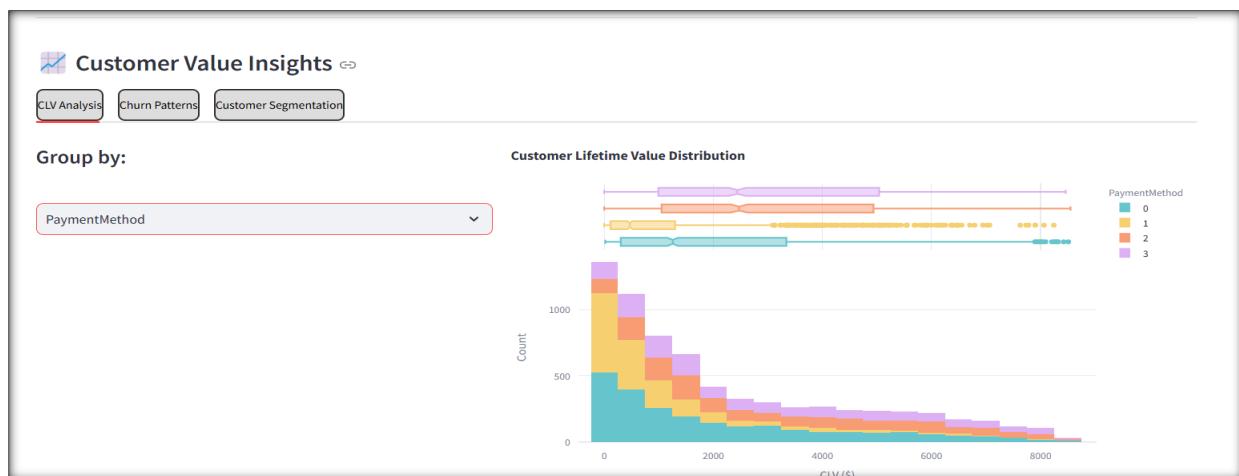
Below are the Charts under this section

	<p>The donut chart displayed represents gender distribution:</p> <ul style="list-style-type: none">• Blue (50.5%) – Represents one gender (likely Male).• Red (49.5%) – Represents the other gender (likely Female).
 <p>Bar chart titled 'Tenure Distribution' showing the number of customers (Y-axis, 0 to 3000) versus tenure in months (X-axis, 0 to 70+). The distribution is skewed, with the highest number of customers having a tenure of 0-6 months, followed by a gradual decline in the mid-tenure range (6-24 months) and a small spike for long-tenure customers (70+ months).</p>	<p>This chart shows the distribution of customer tenure (number of months a customer has stayed with the company).</p> <ul style="list-style-type: none">• Most customers have a low tenure (0-6 months), meaning they leave early.• There is a gradual drop in mid-tenure customers (6-24 months).• A small spike in long-tenure customers suggests some customers stay for a long time.
 <p>Bar chart titled 'Internet Adoption vs Churn' showing the number of customers (Y-axis, 0 to 300) versus service status (X-axis: DSL, Fibre optic, No). The chart includes a color scale legend for 'Churn' ranging from 0% (light blue) to 100% (dark blue). The 'No' category shows the highest churn rate, while 'Fibre optic' shows the lowest.</p>	<p>This chart explores churn rates across different internet service types.</p> <ul style="list-style-type: none">• The bars represent different service adoption levels.• Higher churn for customers without internet services (rightmost bar).• Customers with fibre optic internet tend to churn more than DSL users.



Customer Value Insights

This page provides Customer Lifetime Value (CLV) Analysis, helping businesses understand the long-term revenue potential of their customers.



The top navigation bar includes three main analytical sections:

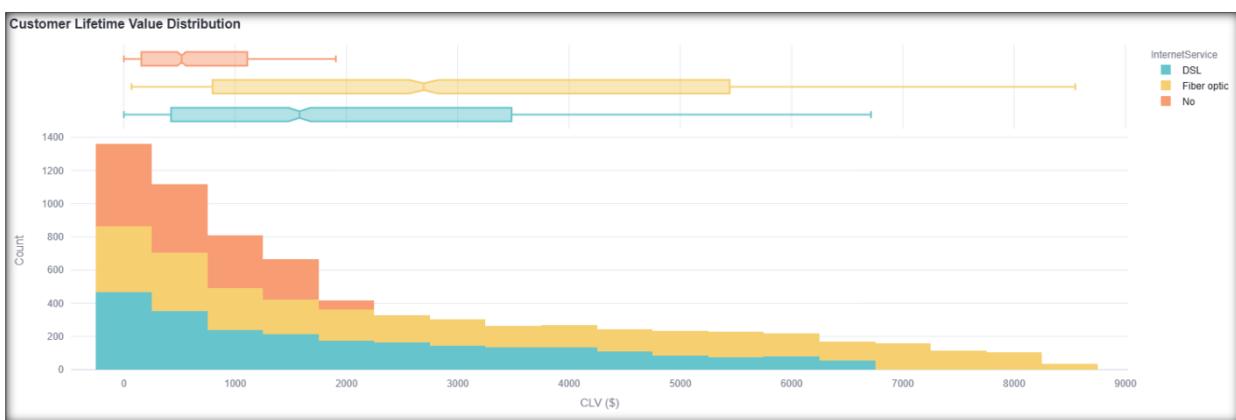
- CLV Analysis (Highlighted): Current view, focusing on CLV distribution.
- Churn Patterns: Likely used for understanding customer retention and attrition.
- Customer Segmentation: Used for categorizing customers based on behaviours and value.

Grouping Dropdown allows segmentation by:

- Contract Type (e.g., Monthly, Yearly)
- Payment Method (Credit Card, Bank Transfer, etc.)
- Internet Service Type (Fiber, DSL, etc.)

Using above grouping below are the sample plots created :-

CLV Analysis

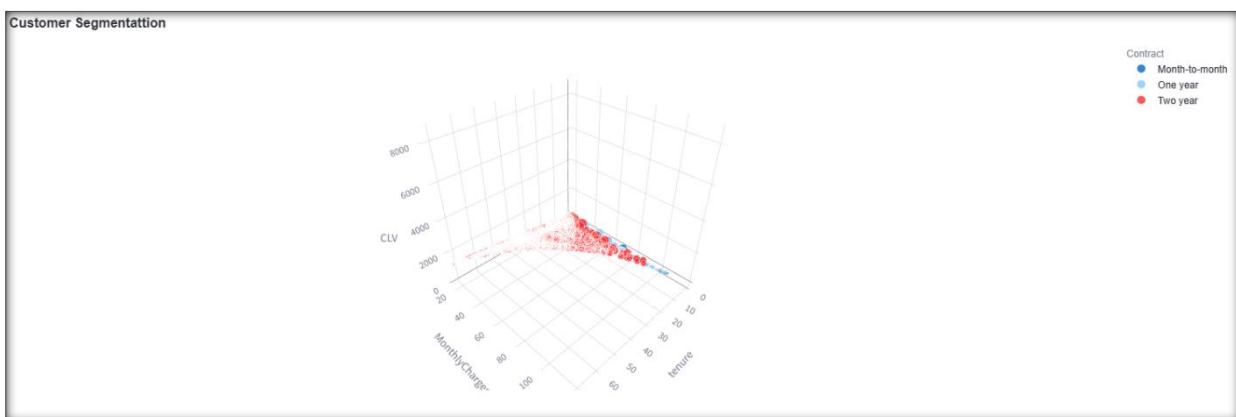


Above visualization shows CLV distribution segmented by Internet Service Type (DSL, Fiber Optic, No Internet Service).

Key Findings:

- DSL Customers (Blue): Show a more balanced spread of CLV, with fewer extreme outliers.
- Fiber Optic Customers (Yellow): Have higher median CLV but a wider range, suggesting premium services drive more revenue but with more variability.
- No Internet Service (Orange): Have the lowest CLV, with a small number of high-value customers.

Customer Segmentation



This 3D scatter plot visualizes Customer Lifetime Value (CLV), Monthly Charges, and Tenure, segmented by contract type:

- Blue (Month-to-Month)
- Light Blue (One-Year Contract)

- Red (Two-Year Contract)

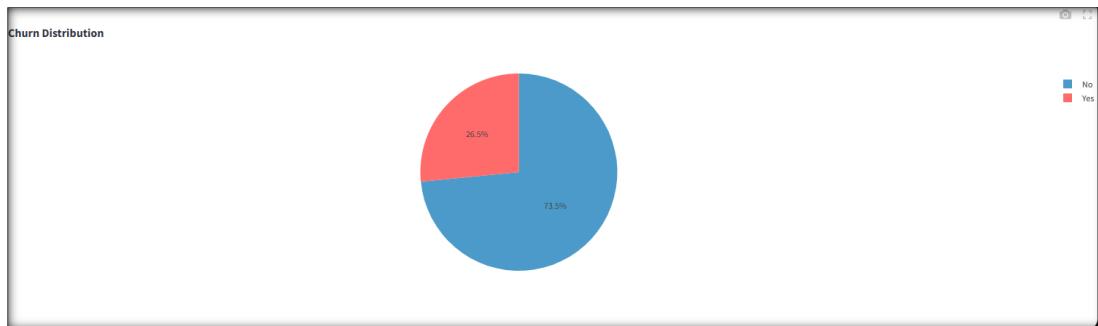
Insights:

- Customers with longer tenure (higher X-axis values) tend to have higher CLV.
- Two-year contract customers (red) dominate the high CLV range, indicating long-term contracts lead to higher lifetime value.
- Month-to-month customers (blue) are concentrated in the lower CLV range, suggesting they churn quickly and contribute less revenue over time.

The **CLV Distribution Graph** shows how much revenue a customer generates before they leave.

CLV Range	Number of Customers	Churn Likelihood
\$0 - \$500	High	Very High (short tenure, low monthly spending)
\$500 - \$2,000	Moderate	Moderate (medium tenure, average spending)
Above \$2,000	Low	Low (long tenure, loyal customers)

Churn Patterns



Above pie chart represents the churn distribution among customers, indicating the proportion of customers who have left (churned) versus those who have remained.

Key Insights:

- 73.5% (Blue) - Retained Customers
 - Majority of customers continue using the service.
 - Indicates a relatively strong customer retention rate.
- 26.5% (Red) - Churned Customers
 - A significant portion of customers have left.
 - Identifying reasons for churn (e.g., pricing, service issues) is crucial for improvement.

Key Insights from the Dashboard

Contract Type vs. CLV:

- Two-year contract customers (red) have the highest CLV, indicating long-term customers bring more revenue.
- Month-to-month customers (blue) have lower CLV, suggesting a higher churn risk and lower long-term revenue contribution.
- One-year contract customers (light blue) are in between, with moderate CLV and retention.

Internet Service Type vs. CLV:

- Fiber Optic customers have the highest CLV but show high variability.
- DSL customers have a stable but lower CLV.
- Customers with no internet service have the lowest CLV, meaning they contribute the least

to long-term revenue.

Churn Analysis:

- 26.5% of customers have churned, which is a significant loss in revenue.
- Majority (73.5%) are retained, but further reducing churn will drive higher profitability.

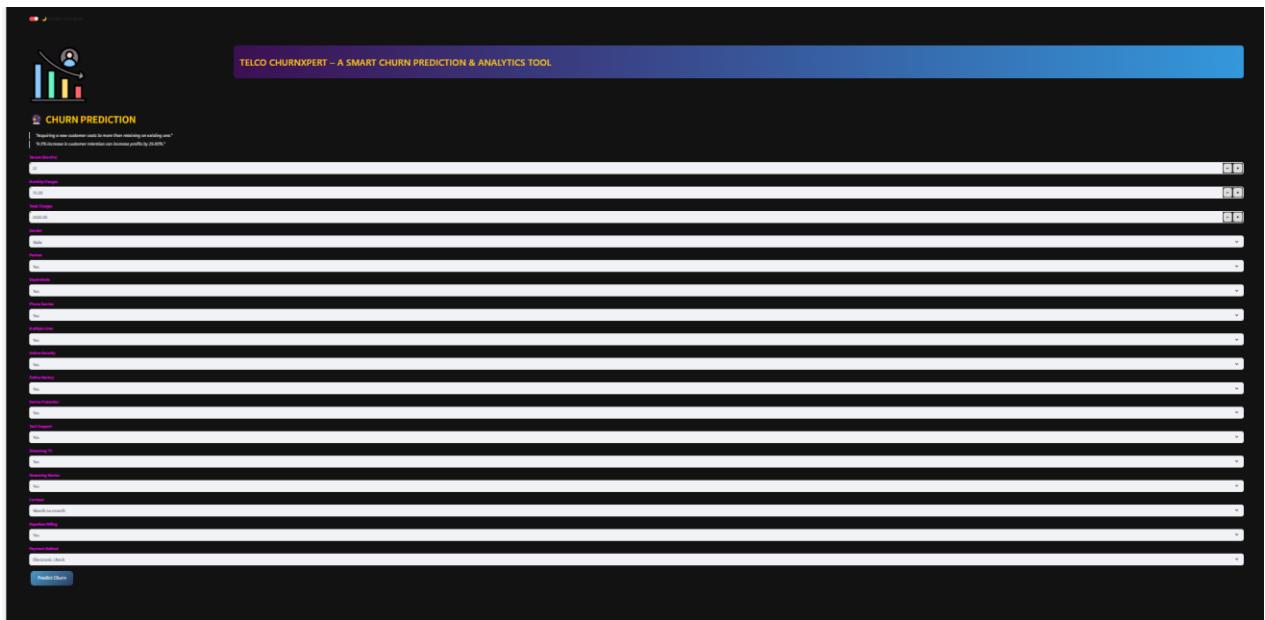
Page3 :Churn Explore and Predict

Churn Prediction Page is a critical part of the **Telco ChurnXpert** application, designed to predict whether a customer is likely to **churn (leave the service provider) or remain subscribed**.

This page allows users (such as customer service teams, business analysts, and marketing managers) to input customer details and get a real-time churn prediction.

By analysing customer demographics, service subscriptions, and billing information, businesses can:

- Identify high-risk customers and proactively offer retention strategies.
- Optimize marketing efforts by targeting at-risk customers with discounts, loyalty programs, or personalized offers.
- Enhance customer support by identifying customers likely to churn and addressing their concerns.



Functionalities of this Page

- Input Form: Users manually enter customer details via input fields (text boxes, sliders, dropdowns).
- Prediction Mechanism: Once the details are submitted, a machine learning model (Random Forest) predicts whether the customer will churn or stay.
- Backend Processing:
 - The data from input fields is fed into a trained churn prediction model.
 - The model analyses the input and returns a probability score.
 - The system displays results in a user-friendly format (e.g., "High Churn Risk" or "Low Churn Risk").

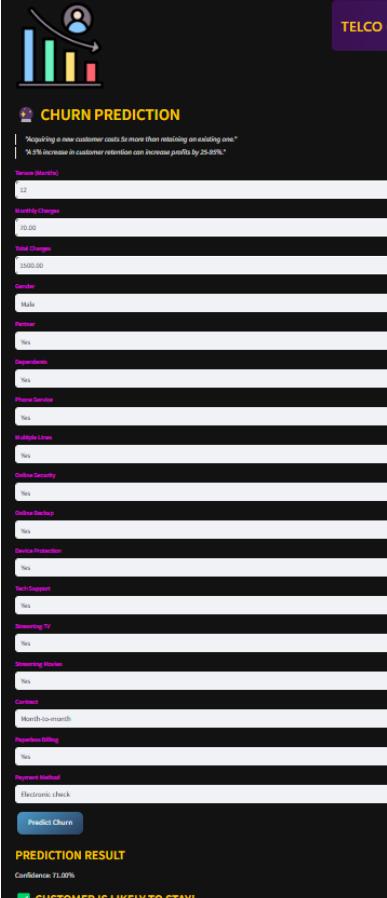
Response of the Page after entering all the details :-

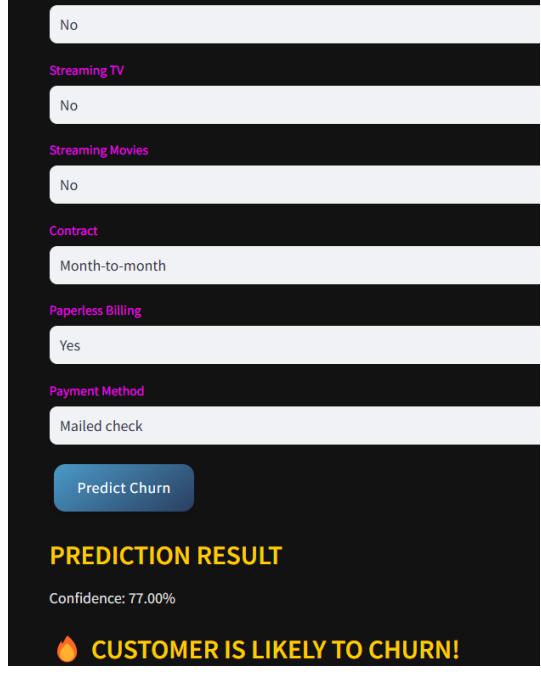
Inputs:

- Customer details such as tenure, monthly charges, contract type, and services used.
- Key factors affecting churn are analysed.

Prediction Process:

- User inputs data and clicks "Predict Churn".
- The model analyses patterns and provides a churn likelihood score.

	<p>Customer is Likely to stay :-</p> <p>Output & Insights: Displays churn probability with a confidence score.</p> <p>Example:</p> <ul style="list-style-type: none">• Prediction: <input checked="" type="checkbox"/> Customer is likely to stay.• Confidence: 71% (high accuracy). <p>Business Impact:</p> <ul style="list-style-type: none">• Helps companies reduce churn by targeting at-risk customers.• Enables data-driven retention strategies (e.g., contract upgrades, personalized offers).
--	---



Customer is Likely to Churn-

Prediction:

- The model predicts that this customer is likely to churn (leave the service) with 77% confidence.

Possible Reasons for Churn Prediction:

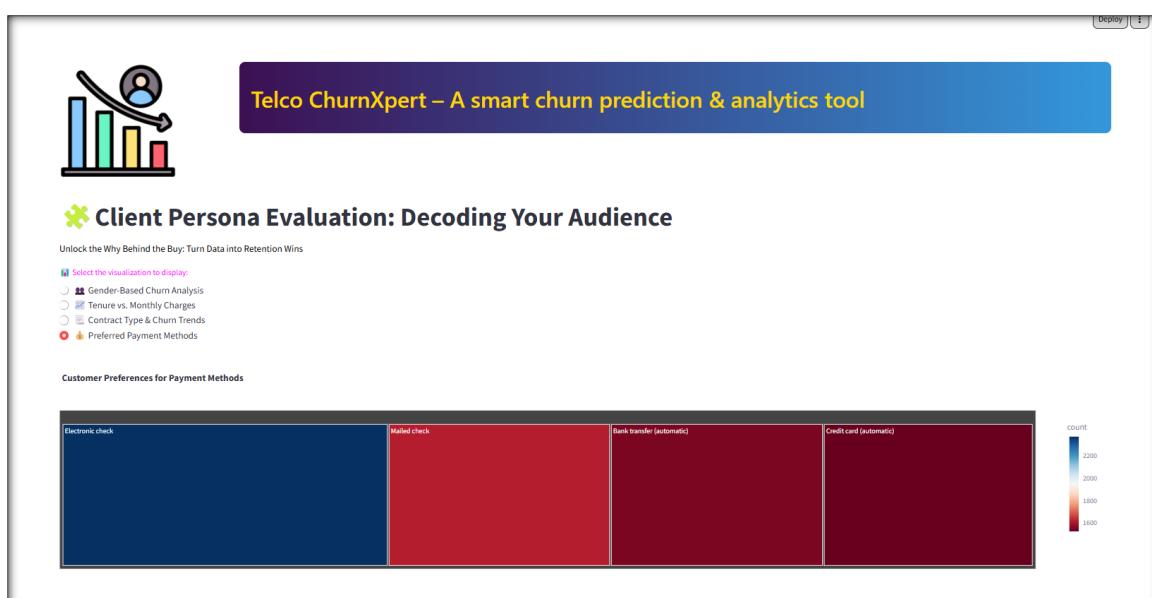
- Month-to-month contract: Customers on flexible contracts are more likely to leave compared to those on long-term contracts.
- Mailed check payment method: Customers using traditional payment methods may have lower engagement with digital services.
- No streaming services: The lack of streaming subscriptions might indicate low engagement with the service.
- Paperless billing: While convenient, it may also indicate customers who are less committed to the service.

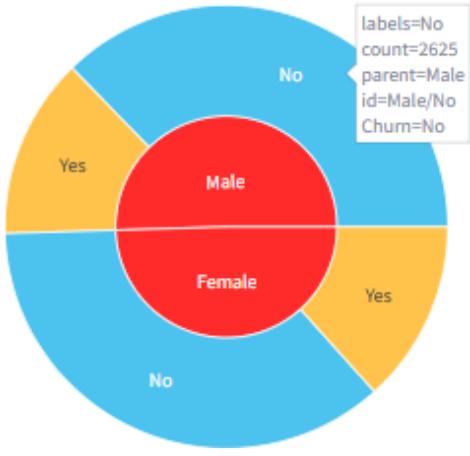
Page4 :Client Persona – Customer Profile Analysis

This page provides a data-driven analysis of customer churn behavior, highlighting key factors influencing retention and attrition. Using insights derived from customer data, the page presents critical takeaways that can guide business decisions for improving customer retention strategies. Each data point is visualized through interactive charts, offering a clear understanding of churn trends and customer behavior. The insights focus on:

- Identifying high-risk customer segments.
- Understanding the impact of pricing, contract types, and additional services.
- Evaluating the role of payment methods and customer engagement in churn.
- Leveraging data-driven strategies to enhance retention and optimize revenue.

As you can see it has Options to choose between Various Graphs -Giving detailed insight





Visualizing the relationship between gender and churn behavior.

Inner Circle (Red Section) → Represents Gender

- The two major sections labeled Male and Female indicate the distribution of customers by gender.

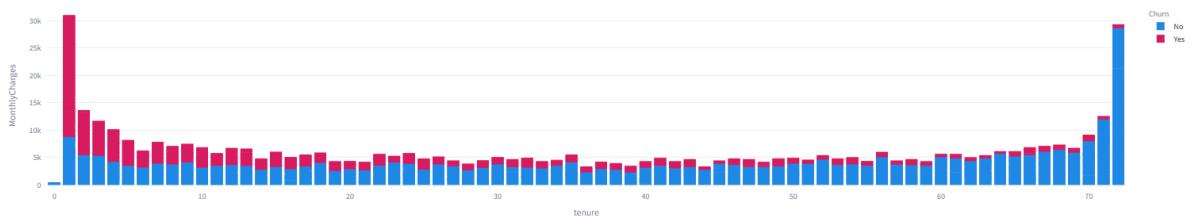
Middle Ring (Green & Orange Sections) → Represents Churn Status

- Each gender category is further divided into Churn = Yes (Orange) and Churn = No (Green).
- Green (No): Customers who have not churned.
- Orange (Yes): Customers who have churned.

Outer Ring Label: "No" (Indicating that this section represents customers who have not churned).

- Count = 2625 (Number of customers in this category).
- Parent = Male (These are male customers).
- Churn = No (They have not churned).

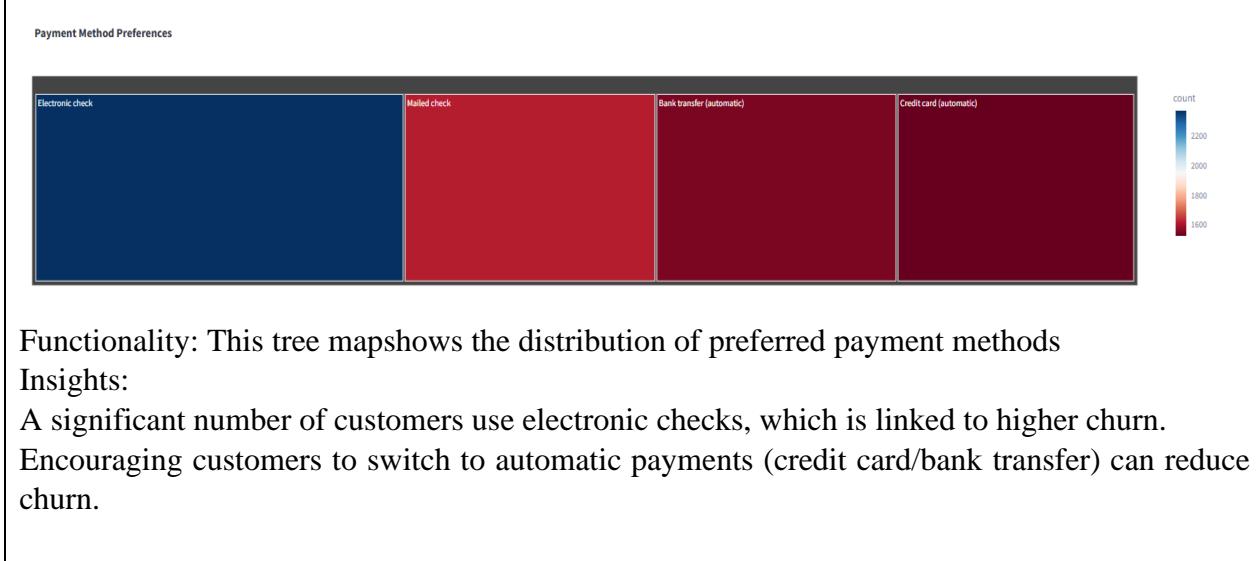
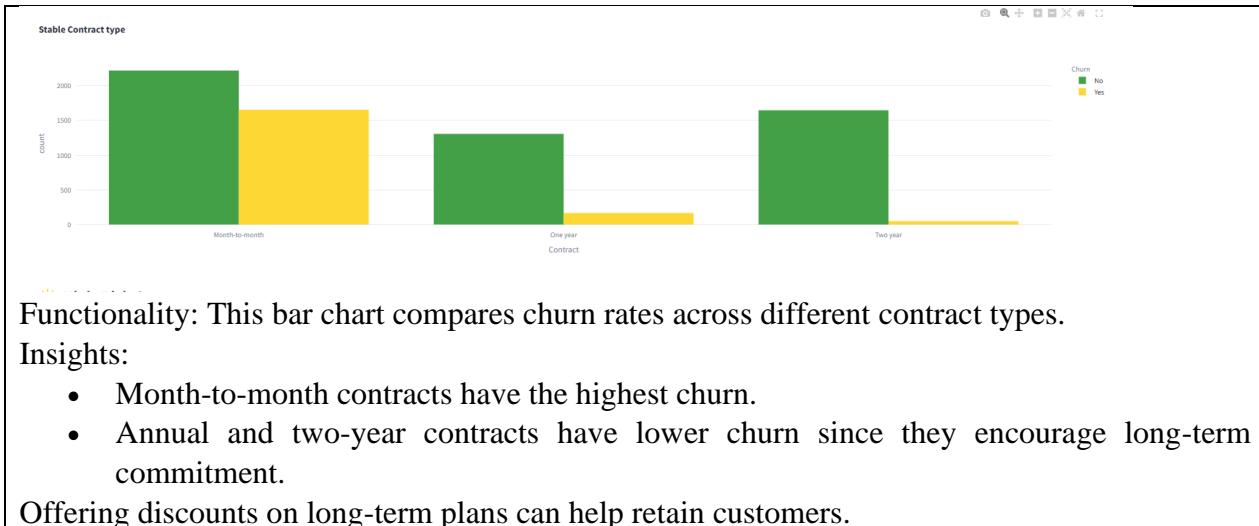
Tenure vs Monthly Charges: Who is at Risk?



Functionality: This **above plot** visualizes how tenure and monthly charges impact churn.

Insights:

- Customers with **low tenure (less than 6 months)** and **high monthly charges** are more likely to churn.
- Businesses should **incentivize early retention** through onboarding discounts or loyalty programs.



Identifying & Retaining High-Risk Customers

💡 High-Risk Customers

customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	OnlineBackup	DeviceProtection	TechSupport	StreamingTV	StreamingMovies	Contract	PaperlessBilling	PaymentMethod	
4	9237-HQITU	Female	0	No	No	2	Yes	No	Fiber optic	No	No	No	No	No	No	Month-to-month	Yes	Electronic check
31	4929-XHJHW	Male	1	Yes	No	2	Yes	No	Fiber optic	No	No	Yes	No	Yes	Yes	Month-to-month	Yes	Credit card (automatic)
47	7760-OPYDY	Female	0	No	No	2	Yes	No	Fiber optic	No	No	No	No	Yes	No	Month-to-month	Yes	Electronic check
65	5122-CYFXA	Female	0	No	No	3	Yes	No	DSL	No	Yes	No	Yes	Yes	Yes	Month-to-month	Yes	Electronic check
80	5919-TMRGD	Female	0	No	Yes	1	Yes	No	Fiber optic	No	No	No	Yes	No	No	Month-to-month	Yes	Electronic check
91	2424-WHJPL	Male	1	No	No	1	Yes	No	Fiber optic	No	No	No	Yes	No	No	Month-to-month	No	Electronic check
115	3071-VBYPO	Male	0	Yes	Yes	3	Yes	No	Fiber optic	Yes	Yes	No	No	Yes	No	Month-to-month	No	Electronic check
122	0404-SWVNG	Male	0	No	No	3	Yes	Yes	Fiber optic	No	No	No	No	No	No	Month-to-month	Yes	Electronic check
139	0390-DCFQD	Female	1	Yes	No	1	Yes	No	Fiber optic	No	No	No	No	No	No	Month-to-month	Yes	Mailed check
166	3376-BMGFE	Female	0	No	No	4	Yes	No	Fiber optic	No	No	No	No	No	No	Month-to-month	Yes	Bank transfer (automatic)

💡 Download High-Risk Customers

📌 Retention Strategy:

- Month-to-month contract users churn the most, requiring incentives for long-term plans.
- High monthly charges increase churn, highlighting the need for tiered pricing and bundled offers.
- Customers without tech support and security services are more likely to leave, making upselling crucial.
- Electronic check users have the highest churn rate, necessitating a push for auto-payment adoption.
- Longer tenure and higher total charges reduce churn, emphasizing early engagement strategies.
- Customer Lifetime Value (CLV) distribution is uneven, requiring targeted retention for high-value users.
- Gender and senior citizen status have minimal impact on churn, shifting focus to behaviour-driven segmentation.
- Leverage these insights to enhance retention strategies and minimize churn!

This page focuses on identifying customers at high risk of churn and providing actionable strategies to improve retention. The top section displays a data table listing customers who are most likely to leave based on key indicators such as contract type, payment method, tenure, and service subscriptions. Below the table, a structured Retention Strategy section provides insights and recommendations for mitigating churn.

Also, we have option to Download the data for Further Analysis

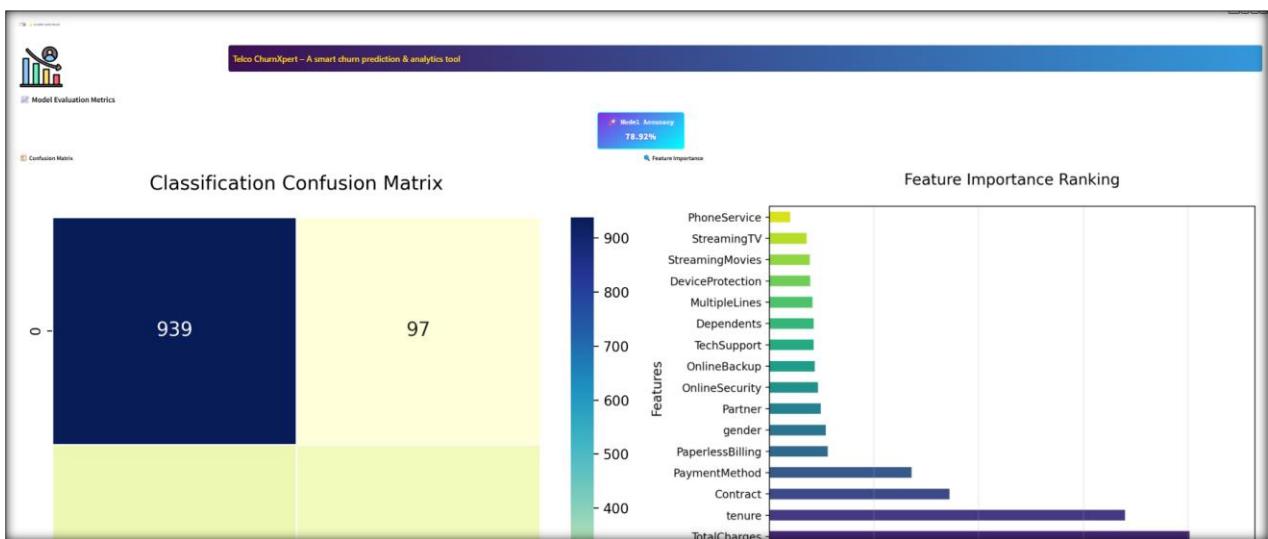
High-Risk Customers Table

Functionality:

- Displays a filtered list of customers predicted to have a high likelihood of churn.
- Key columns include tenure, contract type, internet service type, tech support availability, payment method, and security features.
- A download button allows exporting this data for further analysis or targeted retention efforts.

Page 5 : Model Evaluation

This page provides details on Model evaluation, Accuracy correlation , feature importance analysis, Correlation Heatmap and Confusion Matrix.



A Random Forest model was selected as the churn prediction model because it provides high accuracy, robustness, and interpretability compared to other classification algorithms.

It was trained using below Features such after doing Features important analysis :-

- **Features:** tenure, MonthlyCharges, TotalCharges, Contract, PaymentMethod, Tech Support, etc
- **Target Variable:** Churn (0 for No, 1 for Yes)
- The dataset was split into **80% training** and **20% testing**.

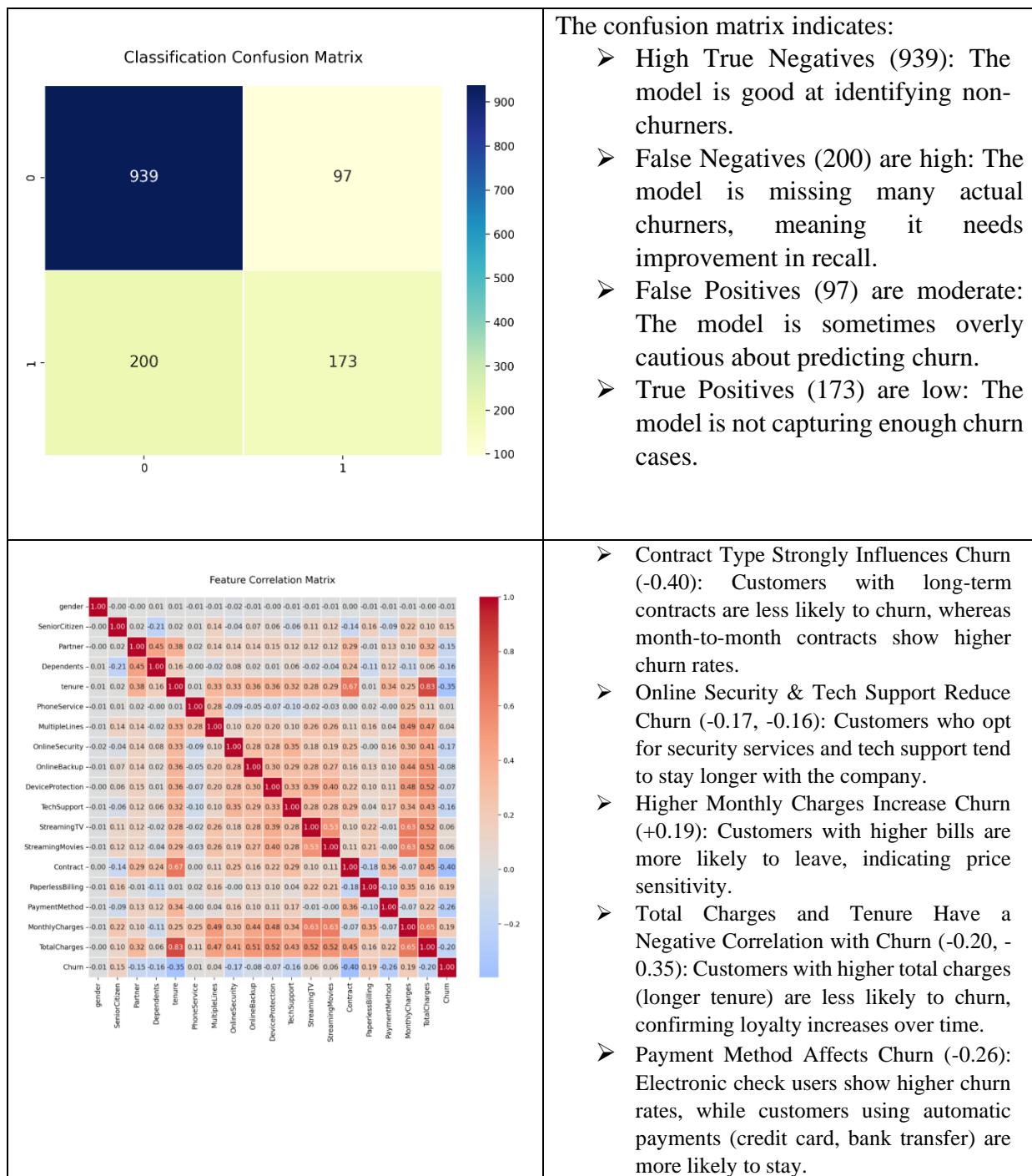
Metric	Score
Accuracy	79%
Precision	64%
Recall	46%
F1-Score	54%

The model correctly predicts churn/no churn 79% of the time. Accuracy = $(TP + TN) / (\text{Total Predictions})$. This is good.

When the model predicts a customer will churn, it is correct 64% of the time. Precision = $TP / (TP + FP)$. A low precision means many false positives (customers predicted to churn but actually stayed).

Out of all actual churned customers, the model correctly identifies 46% of them. Recall = $TP / (TP + FN)$.

The harmonic mean of Precision & Recall. $F1\text{-Score} = (2 \times \text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$. It balances both metrics and is useful when data is imbalanced.



Model Insight :-

Insights

- Month-to-month contract users churn the most, requiring incentives for long-term plans.
- High monthly charges increase churn, highlighting the need for tiered pricing and bundled offers.
- Customers without tech support and security services are more likely to leave, making upselling crucial.
- Electronic check users have the highest churn rate, necessitating a push for auto-payment adoption.
- Longer tenure and higher total charges reduce churn, emphasizing early engagement strategies.
- Customer Lifetime Value (CLV) distribution is uneven, requiring targeted retention for high-value users.
- Gender and senior citizen status have minimal impact on churn, shifting focus to behaviour-driven segmentation.

Recommendations

- Encourage long-term contracts with discounts and exclusive benefits.
- Optimize pricing and introduce bundled discounts to retain cost-sensitive customers.
- Upsell security and support services to improve retention and engagement.
- Promote auto-pay methods by offering incentives for credit card and direct bank payments.
- Identify and engage high-risk customers early with personalized retention offers.
- Focus on maximizing CLV by rewarding loyal, high-value customers.
- Shift from demographic-based targeting to behavioural segmentation for better churn prediction.

Answer 2:

Q2A a) Identify a data-related challenge in your organization that could benefit from the techniques learned in this course (10 marks)

Note- Capgemini is a services company, therefore the answers cover two aspects – first the challenges that Capgemini faces as an organization is covered under the problem statement. Then against each of these challenges a parallel is drawn across various sectors and customers of Capgemini. The data is researched from publicly available information on Capgemini and for some the example stated I have worked as Presales Manager for Data Analytics related Proposal to deliver the solution along with Key Solution and Data Architect

Capgemini is a global leader in consulting, technology services, and digital transformation. Founded in 1967 by Serge Kampf, the company is headquartered in Paris, France. As of 2024, Capgemini employs over 341,100 people across more than 50 countries and generated a revenue of €22.096 billion in 2024. Capgemini has grown significantly through strategic acquisitions and mergers, including notable acquisitions such as Ernst & Young Consulting in 2000, Kanbay for \$1.2 billion in 2007, iGATE for \$4 billion in 2015, Altran in 2019, and Syniti in 2024.

Due to its complex structure, numerous acquisitions, and decentralized operations, Capgemini faces several data-related challenges. Data often resides in isolated systems across different departments, acquisitions, and geographies, leading to silos and fragmentation that hinder comprehensive analysis and decision-making. The lack of standardized reporting frameworks makes performance

tracking and decision-making inefficient across business units. Legacy systems, data fragmentation, and processing inefficiencies slow down analytics, impacting real-time insights and business agility. Inconsistent governance, disparate sources, and manual entry errors result in missing or inaccurate data, affecting compliance and reporting. Managing multiple data systems from acquired firms while ensuring a unified data architecture remains a critical issue.

Personal Contribution to Client's Data Driven Solution :-

I've actively integrated key methodologies into my presales engagements to address client challenges. By advocating modern solutions like cloud-based data lakes, real-time analytics, and AI-driven governance frameworks, I've enabled clients to unify disparate data sources, automate decision-making, and comply with regulations like GDPR and HIPAA.

For instance, a banking client in Germany achieved a 30% faster loan approval process through a centralized data platform, while a U.S. healthcare provider reduced compliance risks by implementing robust data lineage tracking.

As a Presales Manager, my focus lies in bridging technical capabilities with business outcomes. I prioritize:

- Scalable architectures: Demonstrating ROI through proof-of-concepts (e.g., migrating on-premise ERP systems to Azure/AWS for a manufacturing client).
- Stakeholder alignment: Translating complex analytics into actionable insights for C-suite audiences.
- Vendor collaboration: Partnering with tools like Snowflake, Tableau, and Databricks to tailor solutions that align with client maturity levels.

In my role as a Presales **-Data and Solution Architect**, I directly contributed to addressing these challenges by collaborating on RFPs (Request for Proposals) for clients like BSH, Bayer, and global insurers. By combining strategic advisory with hands-on technical validation, I've supported over 15 multi-million-dollar deals in the last two years, driving digital transformation for Fortune 500 clients.

For instance:

- **Designing Data Transformation Strategies:** I mapped client pain points (e.g., fragmented supplier data at BSH) to tailored solutions, such as MDM platforms and AI-driven automation, ensuring alignment with their operational goals.
- **Solution Architecture:** Developed scalable architectures for hybrid cloud integration (e.g., AWS/Oracle Cloud) to unify legacy systems and enable real-time analytics.
- **Stakeholder Collaboration:** Worked with cross-functional teams (data engineers, compliance officers) to define governance frameworks and metrics for ROI measurement.
- **Proposal Development:** Crafted technical and commercial proposals highlighting cost savings (e.g., \$6M saved for an engineering firm through ERP consolidation) and efficiency gains (40% faster supplier onboarding).

Capgemini faces numerous **data-related challenges** due to its **complex structure, acquisitions, and decentralized operations**, as mentioned in the table below :

Challenges	Problems	Sector Examples with Links
Data Silos & Fragmentation	Data exists in isolated systems across different departments, acquisitions, and geographies, limiting seamless interoperability and integration. Capgemini, with its vast global operations and numerous services lines, faces challenges in data silos and fragmentation across its	In the manufacturing industry, companies with complex organizational structures frequently encounter challenges with data silos. For instance, BSH Hausgeräte GmbH, a leading home appliance manufacturer, faced

	<p>diverse business units, acquisitions, and internal systems. Integrating data across multiple geographies, service offerings, and newly acquired companies like Altran requires strong governance, seamless interoperability, and cloud-based solutions. The company has been actively addressing this through Collaborative Data Ecosystems, AI-driven data platforms, and cloud transformation initiatives to unify and streamline internal data flows. However, as it continues expanding organically and inorganically, maintaining a centralized, real-time, and accessible data infrastructure remains an ongoing challenge. Capgemini's investments in AI, automation, and cross-functional data governance will be crucial to overcoming these data fragmentation issues while ensuring operational efficiency and innovation.</p>	<p>difficulties in ensuring data quality and availability across multiple regions and systems. This fragmentation impeded their ability to bring new offerings to market swiftly. To address this, BSH partnered with Capgemini to enhance their master data management, aiming to create a transparent data environment accessible to the entire organization.</p> <p>Capgemini</p>
Reporting Limitations	<p>The lack of standardized reporting frameworks makes performance tracking and decision-making inefficient across business units. Capgemini faces challenges in limited reporting capabilities due to the complexity of managing vast amounts of data across multiple business units, service lines, and acquired entities. The lack of standardized, real-time, and integrated reporting frameworks can hinder decision-making, performance tracking, and operational efficiency. As the company expands, ensuring consistent, scalable, and automated reporting across global teams remains a critical challenge, impacting transparency and data-driven strategy execution.</p>	<p>In the insurance industry, a leading global insurer faced challenges in managing complex Environmental, Social, and Governance (ESG) data sourced from over 30 different channels. This complexity exposed the company to significant audit and reputational risks due to potential inaccuracies in sustainability reporting. To address this, Capgemini collaborated with the insurer to deploy a machine learning-based anomaly detection solution. This innovative approach enhanced data accuracy, improved operational efficiency by 50%, and strengthened compliance in ESG reporting. The successful implementation not only bolstered the insurer's reporting capabilities but also ensured a robust foundation for future sustainability initiatives.</p> <p>Capgemini</p>
Slow Analytics	<p>Legacy systems, data fragmentation, and processing inefficiencies slow down analytics, impacting real-time insights and business agility. Capgemini faces challenges in slow analytics due to the increasing volume and complexity of data generated across its global operations, service lines, and acquisitions. Legacy systems, data fragmentation, and processing inefficiencies can lead to delays in extracting real-time insights, impacting decision-making and client service delivery.</p>	<p>In the manufacturing industry, a prominent company grappled with fragmented data sources, leading to sluggish analytics and decision-making processes. To address this, Capgemini implemented an end-to-end solution that consolidated the company's disparate data into a unified architecture using Microsoft Fabric. This integration not only centralized their data but also resulted in up to a 40% reduction in infrastructure costs and a 30% improvement in performance. The streamlined data processing empowered business units to generate reports and dashboards independently, significantly enhancing agility and responsiveness.</p> <p>Microsoft Partner Network</p>

Data Quality Issues	<p>Data Quality Issues: Inconsistent governance, disparate sources, and manual entry errors result in missing or inaccurate data, affecting compliance and reporting. Capgemini faces challenges in poor data quality and missing data due to the vast and diverse datasets managed across its global operations, service lines, and acquisitions. Inconsistent data governance, disparate data sources, and manual data entry errors can lead to inaccuracies, affecting decision-making, compliance, and reporting.</p>	<p>In the healthcare sector, pharmaceutical companies often rely on historical clinical trial data to inform current research and development. However, much of this valuable data is stored in unstructured formats, such as scanned documents, making it challenging to access and analyze. Bayer Consumer Health faced this exact issue, with critical clinical data trapped in over 3,500 pages of scanned reports. To overcome this, Bayer partnered with Capgemini Engineering to implement an AI-powered data extraction solution. This approach enabled the efficient extraction and standardization of data from complex tables within the documents, transforming previously unusable information into actionable insights. As a result, Bayer could conduct comprehensive meta-analyses without the need for additional clinical trials, accelerating research and reducing costs. Acodis</p>
Data Integration Challenges	<p>Managing multiple data systems from acquired firms while ensuring a unified data architecture remains a critical issue. Capgemini, a global leader in consulting and technology services, has identified data integration as a critical challenge, with 58% of organizations citing it as a primary concern in their data architecture. This challenge is often exacerbated by data silos, reported by 69% of organizations, and data quality issues, affecting 64%.</p>	<p>BMW faced challenges in integrating data across various departments, which impeded efficient decision-making and slowed down processes. To overcome this, BMW collaborated with Capgemini Invent to define and integrate new company processes, derive and track business requirements in an integrated system, and optimize the data structure of the existing system landscape. This initiative led to improved data integration, streamlined Capgemini</p>
Security & Privacy Risks	<p>Cyber threats, regulatory compliance (e.g., GDPR), and decentralized operations create vulnerabilities, requiring strong cybersecurity measures. Capgemini addresses data security and privacy challenges through a comprehensive framework that includes Binding Corporate Rules (BCRs) approved by European Data Protection Authorities. These BCRs ensure a high standard of data protection across all Capgemini entities and facilitate secure international data transfers while complying with GDPR. The company also employs proactive security strategies such as penetration testing, breach simulations, and threat hunting, along with Digital Forensics and Incident Response (DFIR) teams to manage cybersecurity incidents. Despite these measures, Capgemini faced a data breach in 2024, highlighting the need for ongoing vigilance and evolving security practices.</p>	<p>In the financial services sector, a prominent payment technology and software solutions provider sought to migrate its operations to the Google Cloud Platform. The primary concern was to ensure that this transition met stringent security and compliance standards. Collaborating with Capgemini, the company implemented a comprehensive data encryption and protection strategy. This initiative not only facilitated a secure and compliant cloud migration but also enhanced infrastructure scalability and reduced operational costs. The successful deployment enabled near real-time data processing, bolstering the company's ability to respond swiftly to market demands. Capgemini</p>
Lack of Data Governance	<p>Unclear ownership, inconsistent policies, and fragmented compliance structures impede effective data utilization. Capgemini recognizes that data governance is essential for harnessing data effectively and ensuring compliance. In 2020, only 25% of business executives surveyed reported having a complete understanding of their organization's data inventory. To address this, Capgemini has developed a Data Maturity Model that guides organizations through stages from informal practices to innovative, industry-leading data governance. This model emphasizes the importance of formal standards, clear data ownership, and the integration of data governance into business workflows. Additionally, Capgemini offers partnerships with over ten industry-leading vendors, providing consultation, implementation, and operational support to help businesses select the right combination of tools for their data governance needs.</p>	<p>Financial institutions manage vast amounts of sensitive data and must adhere to stringent regulatory requirements. A lack of effective data governance can result in unreliable data for risk aggregation and reporting, potentially leading to non-compliance and financial penalties. Capgemini has addressed this challenge by assisting financial institutions in establishing enforceable data governance frameworks. These frameworks ensure the availability of reliable and accurate data, enhance accountability, and provide traceability, thereby supporting effective risk management and regulatory compliance.</p>

Demonstrating ROI on Digital Investments	Measuring ROI for AI, cloud, and automation projects remains challenging due to unclear financial gains and long-term impact. Capgemini faces challenges in demonstrating ROI for its large-scale technology investments, particularly in cloud, AI, and automation. To address this, the company focuses on data-driven performance tracking, aligning investments with measurable business outcomes, and leveraging internal tools like SAP Fieldglass to improve cost efficiency and workforce productivity. Despite these efforts, quantifying long-term value across diverse service lines remains an ongoing challenge as the company scales its digital transformation initiatives.	In the professional services industry, Capgemini faced the challenge of streamlining operations to reduce operational costs and enhance productivity. To address this, Capgemini implemented Robotic Process Automation (RPA) using the UiPath Platform in conjunction with their Digital Global Enterprise Model (D-GEM) transformation platform. This initiative led to significant improvements, including the clearance of a backlog of 20,000 transactions within a couple of months. The RPA projects demonstrated a substantial ROI, ranging from \$50,000 to \$250,000 per project in annual savings. This clear quantification of benefits facilitated stakeholder buy-in and justified further investments in automation technologies. (uipath.com)
Keeping Up with Rapid Technological Change	The growing volume and variety of data, emerging technologies (AI, quantum computing, blockchain) demand continuous adaptation and workforce upskilling. Capgemini faces challenges in adapting to rapid technological changes and managing the increasing volume and variety of data. The convergence of emerging technologies, such as AI, quantum computing, and biotechnology, requires continuous innovation and integration efforts. Additionally, the exponential growth of data from diverse sources necessitates advanced data management strategies to transform complexity into actionable insights. To address these challenges, Capgemini emphasizes the development of comprehensive data strategies and the adoption of agile operating models that can adapt to the evolving technological landscape.	In the pharmaceutical industry, Boehringer Ingelheim faced the challenge of harnessing vast amounts of unstructured text data to derive actionable insights. To address this, they partnered with Capgemini Invent to develop a custom AI and machine learning-based analytics platform. This platform leverages Natural Language Processing (NLP) algorithms to process previously underutilized data, enabling the company to make more data-driven decisions and stay ahead in a rapidly evolving industry. (Capgemini)
Scalability & Performance	With millions of transactions daily, Capgemini must maintain high-performance data infrastructure while ensuring cost-effective scalability. Capgemini faces challenges in scalability and performance due to the complexity of managing large-scale, innovative projects across diverse industries. A report from the Capgemini Research Institute highlights that while organizations excel at generating innovative ideas, many struggle to scale these innovations effectively, often due to over-reliance on technology and insufficient focus on customer needs.	In the banking and financial services sector, institutions deal with millions of transactions per second. They require high-performance data systems that can handle growth without compromising speed. One of the largest financial institutions globally, based in the UK, supports a diverse clientele with a wide range of financial services and operates a significant global workforce. Their sell-side research platform aimed to improve engagement rates, report readership, and overall site visits while addressing off-platform content consumption. (Source: Capgemini Case Studies)

2b) How Course Learning Can Benefit Finding Solution:

- **Breaking Down Data Silos & Enhancing Integration** Data spread across multiple systems often leads to inefficiencies and incomplete insights.

To address this:

- **Unified Dashboards** in tools like Power BI and Tableau bring all data into a single, easy-to-read interface, helping teams make informed decisions.
- **Standardized Data Models** ensure consistency across departments, reducing redundant efforts and improving collaboration.
- **Automated Data Cleaning** with Python's Pandas library helps merge, clean, and standardize datasets, saving time and ensuring reliability.

- **Integrating External Data Sources** enriches internal datasets, offering a more complete picture of market trends and customer behavior.

➤ **Improving Analytics & Reporting for Better Decision-Making:** -Slow or inadequate reporting can impact how quickly businesses react to market changes.

Solutions include:

- **Advanced Data Visualization** techniques transform raw data into meaningful insights using interactive reports and dashboards.
- **Real-Time Monitoring** ensures key metrics are continuously tracked, allowing quick responses to emerging trends.
- **Predictive Analytics & Anomaly Detection** with machine learning models help businesses anticipate future trends and detect irregularities in their data.
- **Data Storytelling** makes complex information more digestible, allowing stakeholders to understand insights and make strategic decisions confidently.

➤ **Maintaining Data Quality & Strengthening Governance** -Poor data quality can lead to costly mistakes. To maintain data integrity:

Solutions include:

- **Automated Cleaning & Preparation** tools identify and fix errors, handle missing values, and standardize datasets.
- **AI-Powered Anomaly Detection** proactively flags inconsistencies, reducing manual review efforts.
- **Regulatory Compliance & Governance** frameworks ensure businesses meet legal standards, like GDPR, while fostering structured data management.
- **Defined Roles & Responsibilities** in data governance ensure accountability and consistency in data handling.

➤ **Ensuring Data Security & Privacy** With rising cyber threats, organizations must take proactive security measures:

Solutions include:

- **Data Encryption & Key Management** protect sensitive information during storage and transmission.
- **AI-Driven Security Monitoring** detects and responds to potential threats in real-time.
- **Incident Response Systems** allow businesses to act swiftly against security breaches, minimizing damage.

➤ **Demonstrating ROI & Making Data-Driven Investments** -One of the biggest challenges organizations faces is justifying the value of data initiatives.

To prove ROI:

- **SMART Metrics** (specific, measurable, achievable, relevant, time-bound) help track performance and effectiveness.
- **Data-Driven Stakeholder Communication** enhances transparency by using visual reports to showcase performance and financial returns.

➤ **Scaling & Keeping Up with Rapid Technological Advances** As data grows in volume

and complexity, businesses need scalable solutions:

- **Cloud-Based & AI-Optimized Infrastructure** dynamically adjusts resources based on demand, improving efficiency and cost savings.
- **Python-Based ETL Pipelines** automate large-scale data processing, reducing manual work.
- **AI & Neural Networks** enable continuous learning, helping businesses adapt to new trends and innovations.
- **Distributed Computing (e.g., Spark, Cloud ML)** accelerates model training and data processing for improved performance.

2C) Discuss potential implementation challenges and how you would address them

Implementation Challenges and Solutions

Data Silos and Fragmentation

- Multiple systems storing duplicate or inconsistent data.
- Lack of standardized formats and governance policies.
- Difficulties in accessing real-time, integrated insights.

Solution: Implementing an **Intelligent Data Integration Platform (IDIP)** centralizes disparate data sources, leveraging **Robotic Process Automation (RPA)** and **Machine Learning (ML)** for automated data extraction, validation, and cleansing. By using **data mesh architecture**, businesses can decentralize data ownership while maintaining uniformity.

Example: **BSH Hausgeräte GmbH**, a global home appliance manufacturer, faced issues with fragmented supplier data across 50 countries, leading to inefficiencies in procurement and delayed time-to-market. By integrating **Stibo Systems' Multi-Domain Master Data Management (MDM) platform**, they achieved a **360-degree supplier data view**, reducing supplier onboarding time by **40%**, cutting data inconsistencies by **60%**, and enabling faster business decisions.

Reporting Limitations and Data Inconsistencies

- Legacy reporting tools lack real-time capabilities.
- Data inconsistencies across multiple departments.
- Limited predictive insights for decision-making.

Solution: **AI-driven anomaly detection systems** enhance reporting accuracy by identifying and resolving discrepancies in real-time. **Automated compliance reporting** ensures standardized frameworks across global operations. Integrated **data visualization platforms** like Power BI and Tableau provide interactive insights for better decision-making.

Example: A **global insurer** struggled with ESG (Environmental, Social, and Governance) compliance reporting across **30+ regional offices** due to inconsistent data sources. Capgemini deployed a **machine learning-based anomaly detection system**, improving data accuracy by **50%**, reducing manual reporting errors by **75%**, and ensuring compliance with global sustainability mandates.

Slow Analytics Due to Legacy Systems

- High computational loads on aging infrastructure.
- Batch-based data processing delaying insights.

- Inability to support real-time analytics and AI models.

Solution: Transitioning to **cloud-native architectures** on Snowflake, AWS, or Google Cloud eliminates processing bottlenecks. Using **serverless computing and auto-scaling** reduces dependency on on-premises infrastructure. **Federated analytics models** allow organizations to process data closer to its source, reducing latency.

Example: A leading international banking group modernized its **investment research platform** by moving from a **legacy CMS to a cloud-based AEM/React platform**. This increased platform **uptime to 99.98%**, enhanced user engagement by **35%**, and allowed **300+ global analysts** to deliver real-time investment insights, reducing research publication time from **24 hours to under 3 hours**.

Data Quality and Governance Challenges

- Lack of standardized governance policies.
- Inaccurate or incomplete data affecting decision-making.
- Compliance risks with evolving regulatory requirements.

Solution: Establishing a Data Governance Framework ensures data integrity through automated policy enforcement. AI-powered data quality tools detect and correct errors in real-time, while automated metadata management improves transparency. Implementing blockchain-based audit trails enhances compliance reporting.

Example: BSH's transition to a federated master data management system allowed it to eliminate 95% of supplier data duplication issues, reduce compliance risk exposure by 70%, and improve the accuracy of procurement decisions, leading to a €10M annual cost reduction in supply chain inefficiencies.

Integration Complexity of Diverse Data Sources

- Mergers and acquisitions introduce incompatible systems.
- Data stored across multiple cloud and on-premise environments.
- API inconsistencies and lack of standardization.

Solution: API-led integration enables seamless data exchange between systems, supporting event-driven architectures for real-time data flow. Hybrid cloud platforms ensure consistent data governance across multiple environments. Edge computing solutions allow localized data processing to minimize transfer delays.

Example: A global engineering firm integrated Oracle Cloud for supplier payment processing, consolidating 5 different ERP systems into a single integrated platform. This reduced invoice processing time by 30%, improved payment accuracy by 98%, and saved \$6M annually by eliminating duplicate payments.

Security and Compliance Risks

- Increasing cyber threats targeting sensitive business data.
- Compliance complexity with GDPR, CCPA, and other regulations.
- Data sharing challenges across third-party ecosystems.

Solution: Implementing Zero Trust Architecture (ZTA) strengthens security through multi-layered encryption and AI-driven threat detection. Tokenization and homomorphic encryption protect data while enabling analytics without exposure. Federated data governance ensures

regulatory compliance across jurisdictions.

Example: A payment technology provider processing \$50B in transactions annually migrated its customer data security to Google Cloud. By adopting real-time tokenization, the company reduced fraud-related losses by 85%, ensuring GDPR compliance while reducing data storage costs by 40%.

Measuring ROI on Digital Investments

- Difficulty in quantifying benefits from AI and automation.
- Long timeframes for return on investment (ROI) realization.
- Lack of standard metrics to track digital transformation success.

Solution: Developing a comprehensive ROI framework with automated cost-benefit analysis ensures visibility into financial gains. Predictive analytics models estimate long-term impact, while business intelligence dashboards provide real-time investment performance tracking.

Example: Capgemini's ESOAR (Eliminate, Standardize, Optimize, Automate, Robotize) framework reduced manual effort by 60%, increased process efficiency by 45%, and delivered an average 3x ROI within 12 months for enterprises implementing automation across financial operations.

Keeping Up with Rapid Technological Changes

- Constantly evolving AI, blockchain, and quantum computing trends.
- Lack of workforce readiness for new technologies.
- Challenges in integrating emerging tools into legacy environments.

Solution: Establishing an AI-driven digital transformation hub allows organizations to rapidly prototype and deploy new technologies. Continuous learning programs ensure workforce upskilling in AI, cloud, and cybersecurity. Low-code/no-code platforms facilitate faster implementation of advanced technologies.

Example: A global banking firm modernized its financial research division by integrating AI-driven content recommendation engines, increasing customer engagement by 50%, reducing churn by 30%, and improving cross-selling effectiveness by 25%.

Scalability and Performance Optimization

- Increasing data volumes requiring high-performance computing.
- Latency issues affecting real-time analytics.
- Unoptimized resource utilization in cloud environments.

Solution: AI-powered workload orchestration dynamically adjusts computing resources based on demand. Serverless architectures eliminate unnecessary infrastructure costs, while real-time data processing engines enhance query execution speeds. Multi-cloud optimization strategies ensure resilience and cost efficiency.

Example: Capgemini's ESG Cloud with Snowflake helped financial services firms scale their ESG data platforms. This enabled real-time ESG data processing across 100+ compliance jurisdictions, reducing compliance reporting costs by 50%, and cutting ESG data reconciliation time from weeks to just a few hours.

Answer 3:

The XYZ Bank Dashboard provides a comprehensive view of customer spending, credit risk, loan distribution, and financial behavior, enabling data-driven decision-making. It highlights high-risk customers, spending patterns, loan types, and credit card recommendations, helping the bank optimize credit approvals, reduce defaults, and target high-value customers.

Brief documentation explaining your design choices and insights

Dashboard Layout & Structure

To make a simple, intuitive, and easy-to-use dashboard specific to the dataset for XYZ Bank with customer demographics, usage of financial products, expenditure on credit cards, loan portfolio, and investment habits.

The dashboard is aimed at making business leaders and analysts able to promptly review key measures while also being able to dive deeper into some areas of particular interest.

KPIs at the Top: Because banks care most about revenue, customer interaction, and lending performance, the top row of the dashboard showcases major indicators like total revenue, credit card uses, approval rate for loans, and average customer investment.

The dashboard has a systematic flow:

- Customer Overview : Provides Overview
- Customer Demographics & Segmentation: Presents information regarding the target market in terms of age, gender, and geographical location.
- Financial Insights: Displays trends in spending through credit cards, loan approvals, and investments.
- Risk Analysis and Customer Default Prediction: Facilitates an evaluation of repayment behaviour of loans, risks, and revenue-generating potential.
- Credit Card Recommendation Engine : Display Credit History and Credit Loan , Risk Level details.

Choice of Visualizations

- Each chart was specifically selected to best display financial data and customer information:
- KPI Cards: Show key financial indicators, giving an instant snapshot of revenue, transactions, and financial product take-up.
- Line Charts: Monitor time trends, for example, revenue trends, investment habits, and loan authorization rates.
- Bar & Column Charts: Employed when comparing financial product performance, expenditure habits by income group, and loan disbursement.
- Pie Charts: Display the distribution of customers according to demographics, credit card types, and loan categories.

- Heatmaps: Emphasize spending behavior by various credit card levels (Platinum, Gold, Silver, Bronze) to uncover high-value clients.
- Scatter Plots: Demonstrate correlations between salary levels, credit card expenditure, and investment patterns to determine high-net-worth individuals.

Interactivity & Filters - With the varied dataset, filters and slicers enable users to drill down data by customer income, location, product usage, and time period, and thus the dashboard is extremely flexible for varied analyses.

- With the financial nature of the data, interactivity was key to allow for customized analysis:
- Date Filters: Enable the tracking of spending, loan payments, and income across given periods.
- Segment Filters: Facilitate analysis of customer actions in terms of income segment, loan segment, or credit card category.
- Geographical Filters: Provide performance at the city or branch level, determining high-performing areas and under-penetrated areas that require greater outreach.

DAX Query :-

- Total Customers: `DISTINCTCOUNT(Customer[Customer_ID])` ensures unique customer count.
- Avg Monthly Spendings: `AVERAGE(Customer[Monthly Spendings])` tracks spending patterns.
- High-Value Customers: `COUNTROWS(FILTER(Customer, Customer[Monthly Spendings] > 75000))` segments premium users.
- Customers with High Debt: `COUNTROWS(FILTER(Customer, Customer[Debt] / Customer[Salary] > 0.7))` identifies default risk.
- Loan Type Segmentation: `COUNTROWS(FILTER(Customer, Customer[Salary] > 2500000 && Customer[Loan Type] IN { "Home Loan", "Car Loan" }))` groups income-based loan holders.
- Debt-to-Income Ratio: `DIVIDE(Customer[Debt], Customer[Salary], 0)` prevents calculation errors.
- Spending Category: `SWITCH(TRUE(), conditions...)` classifies low, moderate, and high spenders
- Credit Card Recommendation: Suggests Gold, Silver, or Bronze based on salary and spending.
- High-Risk Customers by Age: `COUNTROWS(FILTER(Customer, Customer[Debt_to_Income_Ratio] > 0.7 && Customer[Age] BETWEEN 38 && 44))` detects risk-prone age groups.
- City Spending Trends: `SUMX(FILTER(Customer, Customer[City]), Customer[Monthly Spendings])` maps spending hotspots.

Insights & Strategic Value

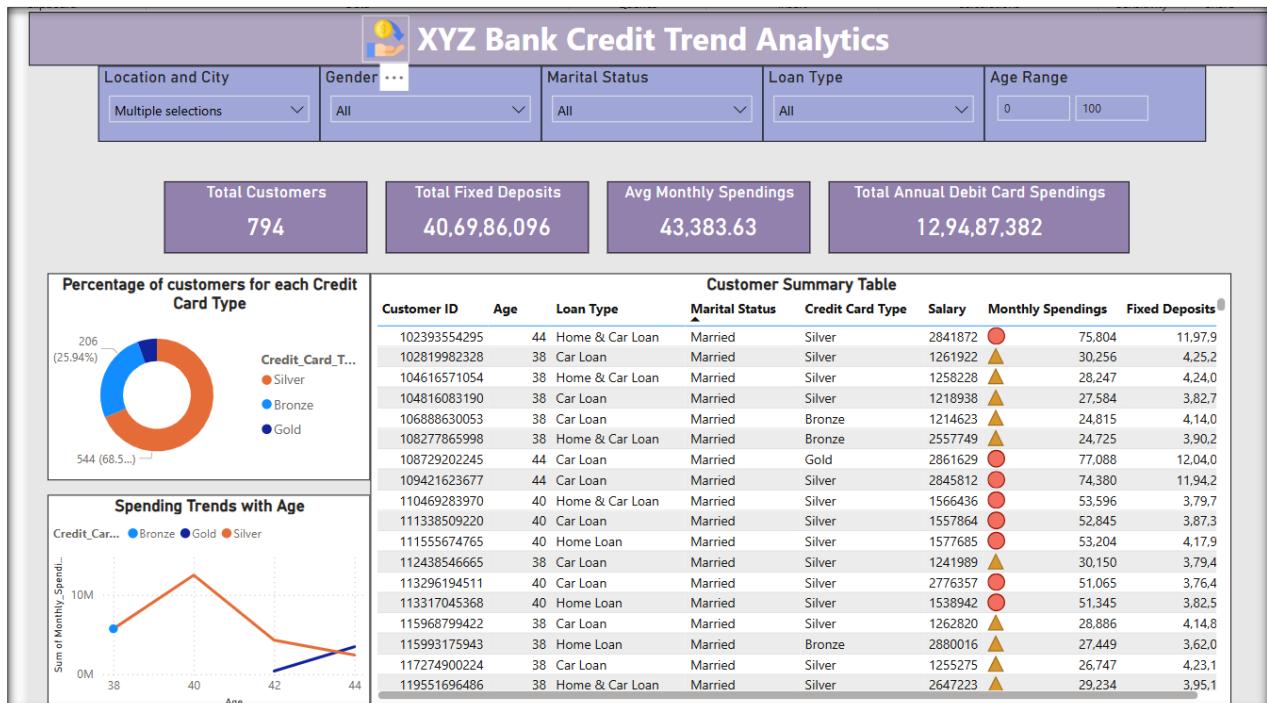
The dashboard interprets raw financial data into business decision-making insights, enabling decision-makers to:

- ✓ Improve Credit Card Targeting:
 - High-value customers prefer Platinum/Gold cards, requiring premium-focused promotions.

- Underutilized credit cards in certain age groups present untapped market opportunities.
- ✓ Maximize Loan Offerings:
 - Home loans dominate; personal loans show steady growth, reflecting evolving financial needs.
 - Detecting delayed repayments improves risk assessment and enables customized repayment plans.
- ✓ Increase Investment Services:
 - Rising mutual fund investments among salaried employees create cross-selling potential.
 - Seasonal equity investment trends highlight the need for adaptive financial advisory services.
- ✓ Customer Segmentation & Engagement:
 - Younger customers (25-35) are prime prospects for long-term financial products.
 - Optimizing retention policies based on recurring transaction patterns.

Screenshots of key dashboard components

- **Overview Dashboard** - The XYZ Bank Credit Trend Analytics dashboard provides a comprehensive view of customer spending, credit card usage, and loan distribution. It highlights that Silver cards dominate (68.5%), peak spending occurs around age 38, and high fixed deposits (₹40.69 Cr) indicate a risk-averse customer base. The insights suggest targeted Gold card promotions, investment advisory services, and loan-linked rewards to enhance customer engagement and revenue



KPI Cards (Top Metrics)

- Total Customers: 794 active customers.
- Total Fixed Deposits: ₹40,69,86,096, indicating high savings behavior.
- Avg Monthly Spendings: ₹43,383.63, showing moderate expenditure.

- Total Annual Debit Card Spendings: ₹12,94,87,382, reflecting strong transaction activity.

Pie Chart - Credit Card Type Distribution

- Silver Cards: 68.5% (544 customers), making it the most used category.
- Bronze Cards: 25.94% (206 customers), indicating an emerging segment.
- Gold Cards: 5.56%, highlighting limited premium card usage.

Line Chart - Spending Trends with Age

- Peak spending occurs around age 38.
- Spending declines after 40 years.
- Gold card usage increases after age 42.
- Silver cards dominate across all ages.

Customer Summary Table

- Customers spending above ₹70,000/month are mostly Silver and Gold cardholders.
- Car Loan holders spend less compared to home loan holders.
- Higher salary does not always result in high fixed deposits.
- Home loan holders have higher savings, balancing debt with deposits.

Summarized Insights & Recommendations

1. Credit Card Upgrade Opportunity: Silver cards dominate (68.5%), but high spenders should be moved to Gold cards.
2. Peak Spending Targeting: The 35-40 age group has the highest spending—target them with promotions.
3. Loan-Linked Offers: Car loan holders spend less; incentivize them with exclusive credit card rewards.
4. Investment Advisory Services: Fixed deposits total ₹40.69 Cr, showing a risk-averse customer base—cross-sell mutual funds.
5. Retention Strategies: High spenders post-40 years shift to savings—offer tailored wealth management plans.

- **Customer Demographic Dashboard** - The Customer Demographic dashboard provides insights into customer spending, loan segmentation, and financial behavior across different cities.



Key Metrics:

- Total Customers: 1,000
- Avg Monthly Spendings: ₹43,728.55
- High-Value Customers: 460
- Customers with High Debt: 341
- Top Investment Customers: 813

Loan Type Segmentation:

- Car Loans are the most common, with a significant share among high-income customers.
- Home & Car Loans and Home Loans are evenly distributed between high-income and upper-middle-income groups.
- Personal Loans have the lowest adoption, suggesting conservative borrowing behavior.

High-Value Spending by City:

- Mumbai, Delhi, and Kolkata contribute the most to high-value transactions.
- Mumbai leads with ₹72.69L in high-income spending, followed by Delhi (₹53.90L) and Kolkata (₹54.52L).
- Chennai & Bengaluru have smaller, yet significant, high-income segments.

Spending Behavior:

- Gold and Silver cardholders show higher monthly spending, particularly those with higher salaries.
- Bronze card users have relatively lower spending, aligning with their income brackets.

Debt-to-Income Ratio:

- Several customers have a high debt-to-income ratio, with some exceeding 85%, indicating potential credit risk.

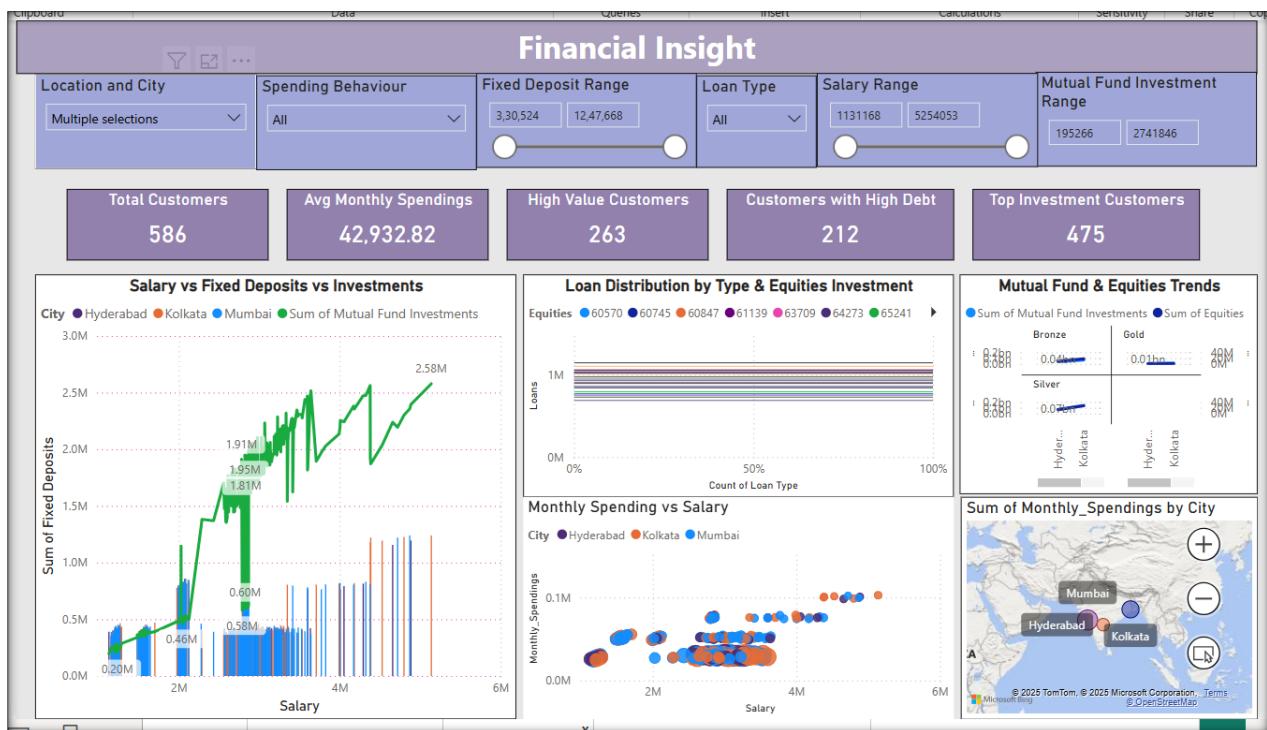
Insights & Recommendations:

- Target Car Loan Customers: Offer cross-selling opportunities for premium credit cards and investment products.
- City-Specific Campaigns: Mumbai, Delhi, and Kolkata have strong high-income

segments—tailor premium financial services for them.

3. Monitor Debt Risk: Customers with high debt-to-income ratios need risk assessment & better repayment plans.
4. Encourage Personal Loans: Introduce customized personal loan offers to balance the loan portfolio.

- **Financial Insights Dashboard:-** The Financial Insight dashboard provides a detailed view of customer investments, loan distribution, spending patterns, and fixed deposits, helping XYZ Bank understand financial behavior and optimize product offerings.



Key Metrics:

- Total Customers: 586
- Avg Monthly Spendings: ₹42,932.82
- High-Value Customers: 263
- Customers with High Debt: 212
- Top Investment Customers: 475

Salary vs Fixed Deposits vs Investments:

- Higher salaries correlate with larger fixed deposits and mutual fund investments.
- Investment activity increases significantly when salaries exceed ₹2M.

Loan Distribution & Equities Investment:

- Loan distribution is evenly spread across equity investment groups.
- Higher equity investments align with larger loan approvals, indicating wealthier customers leverage loans for financial growth.

Mutual Fund & Equities Trends:

- Silver and Gold cardholders invest more in mutual funds and equities than Bronze

cardholders.

- Investment activity is higher in Mumbai and Hyderabad, suggesting city-based targeting opportunities.

Monthly Spending vs Salary:

- Higher salaries drive higher monthly spending, particularly in Mumbai, Kolkata, and Hyderabad.
- Spending plateaus around the ₹4M salary range, indicating saturation.

Geographic Spending Trends:

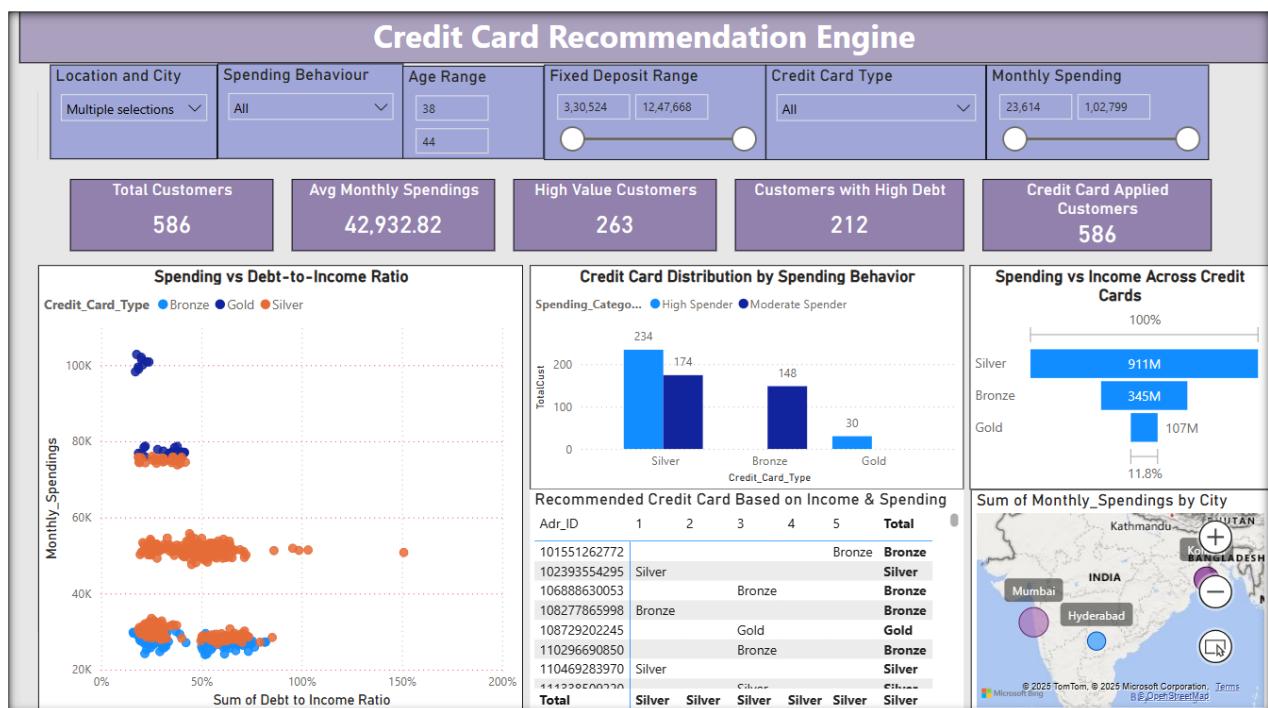
- Mumbai leads in monthly spending, followed by Hyderabad and Kolkata.

Insights & Recommendations:

1. Increase High-Value Investment Services: Leverage fixed deposit holders for investment product cross-selling.
2. Loan-Equity Alignment: Offer customized loan options for high-equity investors.
3. City-Based Financial Planning: Focus investment products in Mumbai & Hyderabad where activity is highest.
4. Spending-Linked Offers: Introduce tailored spending rewards for high-salaried customers nearing spending saturation.

➤ Credit Card Recommendation Engine dashboard

The Credit Card Recommendation Engine dashboard helps XYZ Bank analyze customer spending behavior, debt-to-income ratios, and credit card distribution to provide personalized credit card recommendations.



• Key Metrics:

- Total Customers: 586
- Avg Monthly Spendings: ₹42,932.82
- High-Value Customers: 263

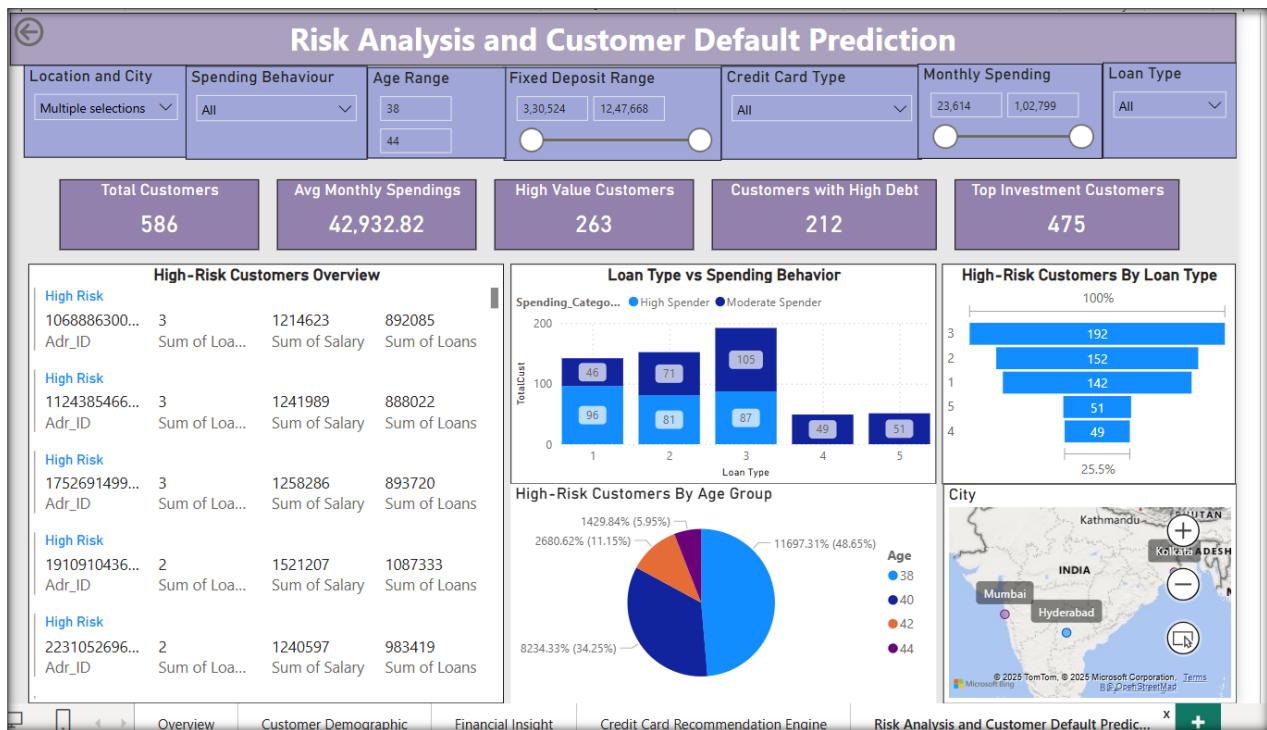
- Customers with High Debt: 212
- Credit Card Applied Customers: 586
- Spending vs Debt-to-Income Ratio:
 - Customers with lower debt-to-income ratios tend to have higher monthly spending.
 - High-debt customers mostly use Silver and Bronze credit cards, with Gold users maintaining better spending balance.
- Credit Card Distribution by Spending Behavior:
 - 234 High Spenders hold Silver Cards, making it the most used category.
 - Bronze Card holders (148) include a mix of high and moderate spenders, showing a transition segment.
 - Gold Cards are underutilized (only 30 users), indicating a missed opportunity for premium upgrades.
- Spending vs Income Across Credit Cards:
 - Silver Cards contribute ₹911M in spending, making it the dominant category.
 - Bronze Cards contribute ₹345M, indicating significant usage among mid-tier customers.
 - Gold Card spending is only ₹107M (11.8% of total), signaling low adoption.
- Credit Card Recommendations Based on Income & Spending:
 - Customers are assigned card upgrades based on their financial behavior, ensuring better credit card utilization.
 - Many Silver cardholders qualify for Gold cards, indicating potential upselling opportunities.
- Geographic Spending Trends:
 - Mumbai and Hyderabad have the highest spending concentrations, making them key cities for credit card promotions.

Insights & Recommendations:

1. Gold Card Upsell Opportunity: Many high spenders are still using Silver cards—target them with premium upgrades.
2. Risk-Based Credit Limit Adjustments: Customers with high debt-to-income ratios should have controlled credit increases.
3. Bronze to Silver Conversions: Encourage Bronze cardholders with moderate spending to move to Silver cards through targeted promotions.
4. City-Specific Credit Campaigns: Mumbai & Hyderabad have high spending activity—run localized card promotions.

➤ **Risk Analysis and Customer Default Prediction dashboard –**

The Risk Analysis and Customer Default Prediction dashboard provides insights into high-risk customers, their loan types, spending behavior, and geographical distribution, helping XYZ Bank mitigate financial risk and prevent defaults



Key Metrics:

- Total Customers: 586
- Avg Monthly Spendings: ₹42,932.82
- High-Value Customers: 263
- Customers with High Debt: 212
- Top Investment Customers: 475

High-Risk Customers Overview

- Lists customers at high risk of default with details of their salary and outstanding loan amounts.
- High-risk customers have large outstanding loans relative to their salaries, making them vulnerable to default.

Loan Type vs Spending Behavior

- Loan holders are classified as High Spenders and Moderate Spenders.
- Customers with higher spending behavior tend to have loans spread across multiple types.
- A significant portion of high spenders have large loan amounts, indicating a potential risk factor if financial instability arises.

High-Risk Customers by Loan Type

- Loan Type 3 (Highest risk group) has 192 high-risk customers, followed by Loan Types 2 (152 customers) and 1 (142 customers).
- Customers with these loan types contribute the most to potential default risks, requiring close monitoring.

High-Risk Customers by Age Group

- 48.65% of high-risk customers fall within the 38-year age group, making them the largest segment at risk.
- The 40- and 42-year-old groups also hold a significant portion of high-risk customers, contributing 34.25% and 11.15%, respectively.

- Older age groups (44 years) have lower default risk, likely due to higher financial stability and lower borrowing.

Geographical Distribution of High-Risk Customers

- Mumbai and Hyderabad have the highest concentration of high-risk customers, indicating areas where loan monitoring and risk mitigation strategies should be prioritized.
- Kolkata also has a presence but with lower risk concentration compared to Mumbai and Hyderabad.

Insights & Recommendations:

1. Targeted Risk Management: Focus on Loan Type 3 customers, as they represent the highest risk.
2. Debt Restructuring Programs: Customers aged 38-42 are at high risk—offering restructured payment plans can reduce defaults.
3. City-Based Monitoring: Mumbai and Hyderabad require stricter risk evaluation, as they have the highest default-prone customers.
4. Spending Control on High-Risk Customers: Monitor high spenders with high loan amounts to prevent over-leveraging.

Your insights from the analysis

Conclusion :-

- Customers with loan amounts exceeding 70% of their salary are at high risk of default. Most high-risk customers earn between ₹12L-₹15L annually but have outstanding loans of ₹8.9L+, increasing default risk
- Loan Type 3 (Unsecured Loans) has the highest risk (192 customers), followed by Loan Type 2 (152 customers).
- High spenders tend to have larger loan amounts, making them more vulnerable to financial instability.
- Unsecured loans (Type 3) are more likely to default, requiring priority monitoring. Secured loans (Home, Car) have lower default risk, indicating a more stable customer base.
- 38-year-olds make up 48.65% of high-risk customers, followed by 40-year-olds (34.25%), making younger borrowers more vulnerable.
- Older borrowers (42-44 years) have lower default risks, suggesting better financial stability.
- Mumbai & Hyderabad have the highest concentration of high-risk customers, while Kolkata has lower default rates.
- High-risk loan distribution suggests unsecured loans are more likely to default, requiring stronger risk controls.

XYZ Bank Strategic Recommendations

Risk Management & Loan Oversight

- Tighter Loan Screening: Stricter terms for unsecured loans while promoting secured loans (home/car).
- Flexible Repayment for At-Risk Borrowers: Offer customized plans for 38-40-year-olds with higher default rates.

- Dynamic Credit Limits: Reduce exposure for high-risk borrowers while rewarding reliable customers.
- City-Specific Risk Control: Stricter approvals for Mumbai & Hyderabad based on high-risk trends.
- Risk-Based Pricing: Higher interest rates for high-risk unsecured loans, better terms for secured loans.
- Proactive Engagement: Early warning systems + personalized repayment support for high-risk borrowers.
- Fraud Monitoring: Detect high-spending, high-debt customers to prevent fraud & overleveraging.
- Financial Education: Targeted campaigns to improve repayment discipline and reduce defaults.

Credit Card Strategy

- Premium Targeting: Special promotions for Platinum & Gold cardholders.
- Increase Credit Card Penetration: Encourage usage in underutilized age groups through incentives.

Loan Portfolio Optimization

- Strengthen Home Loan Market: Maintain dominance with steady personal loan growth.
- Early Repayment Monitoring: Detect delays early & adjust repayment plans to reduce defaults.

Investment & Wealth Expansion

- Cross-Sell Mutual Funds: Target salaried employees showing increased investment interest.
- Seasonal Investment Advisory: Align equity recommendations with market trends.

Customer Segmentation & Retention

- Younger Customers (25–35): Promote long-term savings & mortgage plans.
- Retention Strategies: Use recurring transaction patterns to personalize loyalty incentives.

Data-Driven Growth & Efficiency

- AI-Powered Risk Models: Identify default risks early and adjust lending strategies.
- Revenue Growth: Expand high-margin secured loans & investment services.
- Operational Optimization: Improve loan processing efficiency, rewards programs, and advisory services.