Customer Insights Analysis SaaS Web Application

December 1st, 2024

Sharan P

Abstract

This project presents the development of a Customer Insights Analysis SaaS Application designed to tackle the challenges businesses face in understanding customer behavior and optimizing their marketing strategies. Many companies struggle with high customer churn rates, ineffective customer segmentation, and the inability to derive actionable insights from customer feedback. This application offers tools for customer churn rate prediction, customer segmentation, sentiment analysis, and review scoring, enabling businesses to make data-driven decisions. The platform features a fast and robust API, allowing seamless integration of these insights into existing workflows. By providing an accessible solution, particularly for smaller businesses, the application reduces the need for significant infrastructure investments while enhancing overall business performance. The report also addresses technical implementation, key compliance issues, and the positive impact of the application on business outcome.|

1. Problem Statement:

In today's fast-paced market, understanding and connecting with customers can be a daunting challenge for businesses. When companies struggle to analyze customer behavior effectively, they risk missing out on valuable opportunities and could even face significant revenue loss. Without insights into what customers truly want and need, they can end up losing them to competitors, leading to high churn rates that hurt their bottom line. This lack of tailored strategies can lead to generic campaigns that simply don't resonate, resulting in wasted marketing efforts and lower engagement. While businesses gather customer reviews and feedback, they often lack the tools to dive deep into this data. As a result, they miss out on crucial insights that could inform product enhancements and improve customer service. Without the ability to foresee how customers will act, businesses tend to react to trends rather than anticipate them. To address these challenges, the Customer Insights Analysis SaaS Application offers a comprehensive solution. By providing tools for customer churn rate prediction, effective segmentation, sentiment analysis, and review scoring, this application empowers businesses to make data-driven decisions. With its suite of machine learning-powered services, organizations can enhance customer engagement, optimize marketing strategies, and ultimately improve business performance.

2. Market/Customer/Business Need Assessment:

A comprehensive analysis of the current market trends, customer preferences, and the evolving demands of businesses has led to the development of the Customer Insights SaaS Application. This assessment delves into the intricacies of customer behavior and the challenges faced by organizations in understanding and engaging their target audience, aiming to equip businesses with the right tools to build strong relationships with their customers.

2.1 Market Need Assessment:

2.1.1 The Rise of Data-Driven Decision Making: In an era where data is abundant, businesses are increasingly recognizing the importance of leveraging data analytics to drive their decisions. Companies are seeking solutions that not only provide insights into customer behavior but also translate these insights into actionable strategies that enhance customer engagement and loyalty.

2.1.2 The Demand for Customer-Centric Strategies: As competition intensifies across various industries, businesses are prioritizing customer-centric approaches to stand out. Organizations are realizing that understanding their customers' preferences and pain points is essential for delivering personalized experiences that resonate and foster brand loyalty.

2.1.3 The Shift Toward Predictive Analytics: With advancements in technology, there is a growing interest in predictive analytics among businesses looking to anticipate customer needs and trends. Organizations want to move beyond reactive strategies and instead implement measures that can help them stay ahead of the curve and optimize their offerings for better customer satisfaction

2.2 Customer Need Assessment:

2.2.1 Access to Actionable Insights: Customers today are not just looking for data; they want actionable insights that can help them improve their decision-making. They require tools that provide clear, understandable information about customer trends, preferences, and behaviors, allowing them to tailor their strategies accordingly

2.2.2 Desire for Enhanced Customer Experiences: There is a growing expectation among customers for businesses to create personalized and memorable experiences. Consumers want to feel valued and understood, leading them to seek brands that can adapt to their individual needs and provide tailored solutions that enhance their overall experience.

2.2.3 Integration of Feedback Mechanisms: Customers increasingly want to engage in a two-way dialogue with businesses. They seek platforms that enable them to provide feedback easily and expect organizations to listen and act on their suggestions. This desire for communication helps businesses improve their offerings based on real-time customer insights.

2.3 Business Need Assessment:

2.3.1 Demand for Data-Driven Strategies: In an increasingly competitive landscape, businesses recognize the need to adopt data-driven approaches to remain relevant and efficient. They require tools that can analyze customer data effectively, providing insights that inform strategic decisions and drive growth.

2.3.2 Need for Operational Efficiency: Companies are constantly looking for ways to streamline their operations and reduce costs. There is a pressing need for solutions that enhance productivity and automate processes, allowing organizations to allocate resources more effectively and focus on core business functions.

2.3.3 Focus on Customer Retention: As acquiring new customers becomes more challenging and expensive, businesses are prioritizing customer retention strategies.

3. Target Specifications and Characteristics

3.1 Audience Targeted:

3.1.1 Small to Medium-Sized eCommerce Businesses: Businesses using platforms like Shopify, WooCommerce, and BigCommerce that require insights into customer behavior, retention strategies, and sales trends.

3.1.2 Retailers with POS Systems: Businesses operating physical stores that also utilize POS systems for sales, needing analytics to improve customer experience and operational efficiency.

3.1.3 Marketplace Sellers: Sellers on platforms like Amazon, eBay, and Etsy who want to understand customer sentiments and improve their product offerings based on data analysis.

3.1.4 Marketing and Analytics Teams: Teams within organizations that utilize machine learning for customer segmentation and targeted marketing strategies.

3.1.5 Consultants and Data Analysts: Professionals working with eCommerce businesses who need tools to provide insights and recommendations based on data.

3.1.6 Startups in Retail and eCommerce: New businesses looking for affordable, scalable solutions to analyze their customer data and drive growth from the outset.

3.1.7 Subscription based Services: Businesses that provide subscription products.

3.2 Core Functionality and Design:

3.2.1 Full Functionality Web Application

User-Friendly Interface: The application can provide a comprehensive web interface where users can access all features and functionalities without requiring technical expertise. Users can easily navigate through dashboards, visualize data, and generate reports to gain insights.

Integrated Tools: It can include built-in tools for data analysis, such as customer churn prediction models, sentiment analysis modules, and segmentation capabilities, allowing users to perform all tasks within a single platform.

3.2.2 API Access

API Integration: The application can offer a API that allows eCommerce sites and other businesses to integrate machine learning functionalities directly into their existing systems. This means they can programmatically access features such as churn prediction, customer segmentation, and review scoring.

Customization and Flexibility: By providing API access, businesses can customize how they use your services, embedding them into their workflows or applications as needed. This flexibility makes it easier for companies to leverage your machine learning insights alongside their existing data and processes.

3.2.3 Visualization and Reporting

Custom Dashboards: Allow users to create customizable dashboards to visualize key metrics and insights tailored to their specific business needs. This can help users easily track performance indicators.

Automated Reporting: Enable automated report generation that summarizes key findings, trends, and insights. Users can receive these reports via email or within the application at scheduled intervals.

3.2.4 Machine Learning Model Training and Fine-Tuning

Model Customization: Offer users the ability to fine-tune existing machine learning models with their data, allowing them to adapt the models to their specific business context. This ensures that the models are not only accurate but also relevant to the unique challenges and opportunities faced by each business.

AutoML Capabilities: Include automated machine learning features that simplify the process of selecting, training, and optimizing models based on user data.

3.2.5 User Management and Collaboration

Multi-User Access: Provide options for multiple users within an organization to access the application, with customizable permissions and roles to control data access and functionality.

Collaboration Tools: Incorporate features that allow teams to collaborate on insights and strategies within the application, such as commenting on reports or sharing dashboards.

4. External Search(Information source / References)

References:

1. **Why customer churn rate prediction is important?**
2. **Customer churn prediction for customer retention**
3. **How to choose the best prediction model?**

Datasets:

Teclo customer churn dataset

Each row represents a customer, each column contains customer's attributes described on the column Metadata.

1. The data set includes information about:
2. Customers who left within the last month – the column is called Churn
3. Services that each customer has signed up for – phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies
4. Customer account information – how long they've been a customer, contract, payment method, paperless billing, monthly charges, and total charges
5. Demographic info about customers – gender, age range, and if they have partners and dependents

Code Implementation:

1. Importing all the necessary libraries

```
[80] import pandas as pd
     import numpy as np
     import seaborn as sns
     import matplotlib.pyplot as plt
     import missingno as msno
     import plotly.express as px
     import plotly.graph_objects as go
     from plotly.subplots import make_subplots
     import plotly.figure_factory as ff
     from plotly.offline import download_plotlyjs, init_notebook_mode, iplot

     from sklearn.preprocessing import StandardScaler
     from sklearn.model_selection import train_test_split, cross_val_score
     from imblearn.over_sampling import BorderlineSMOTE

     from sklearn.linear_model import LogisticRegression
     from sklearn.svm import SVC
     from sklearn.neighbors import KNeighborsClassifier
     from sklearn.tree import DecisionTreeClassifier

     from sklearn.metrics import confusion_matrix, accuracy_score, classification_report
     from sklearn.model_selection import StratifiedKFold, RandomizedSearchCV
```

2. Uploading the dataset

```
from google.colab import files
uploaded = files.upload()
```

```
Choose Files  Telco Custo...rn Rate.csv
• Telco Customer Churn Rate.csv(text/csv) - 977501 bytes, last modified: 1/5/2023 - 100% done
Saving Telco Customer Churn Rate.csv to Telco Customer Churn Rate.csv
```

3. Printing the first 10 rows

```
df.head(10)
```

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	...	DeviceProtection	TechSupport	StreamingTV	StreamingMovies	Contract	PaperlessBilling	PaymentMetho
0	7590-VHVEG	Female	0	Yes	No	1	No	No phone service	DSL	No	...	No	No	No	No	Month-to-month	Yes	Electroni chec
1	5575-GNVDE	Male	0	No	No	34	Yes	No	DSL	Yes	...	Yes	No	No	No	One year	No	Mailed chec
2	3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	Yes	...	No	No	No	No	Month-to-month	Yes	Mailed chec
3	7795-CFOCW	Male	0	No	No	45	No	No phone service	DSL	Yes	...	Yes	Yes	No	No	One year	No	Bank transfe (automatic
4	9237-HQITU	Female	0	No	No	2	Yes	No	Fiber optic	No	...	No	No	No	No	Month-to-month	Yes	Electroni chec
5	9305-CDSKC	Female	0	No	No	8	Yes	Yes	Fiber optic	No	...	Yes	No	Yes	Yes	Month-to-month	Yes	Electroni chec
6	1452-KIOVK	Male	0	No	Yes	22	Yes	Yes	Fiber optic	No	...	No	No	Yes	No	Month-to-month	Yes	Credit car (automatic
7	6713-OKOMC	Female	0	No	No	10	No	No phone service	DSL	Yes	...	No	No	No	No	Month-to-month	No	Mailed chec
8	7892-POOKP	Female	0	Yes	No	28	Yes	Yes	Fiber optic	No	...	Yes	Yes	Yes	Yes	Month-to-month	Yes	Electroni chec
9	6388-TABGU	Male	0	No	Yes	62	Yes	No	DSL	Yes	...	No	No	No	No	One year	No	Bank transfe (automatic

10 rows × 21 columns

4. Describing the dataset

	SeniorCitizen	tenure	MonthlyCharges
count	7043.000000	7043.000000	7043.000000
mean	0.162147	32.371149	64.761692
std	0.368612	24.559481	30.090047
min	0.000000	0.000000	18.250000
25%	0.000000	9.000000	35.500000
50%	0.000000	29.000000	70.350000
75%	0.000000	55.000000	89.850000
max	1.000000	72.000000	118.750000

5. Exploratory Data Analysis

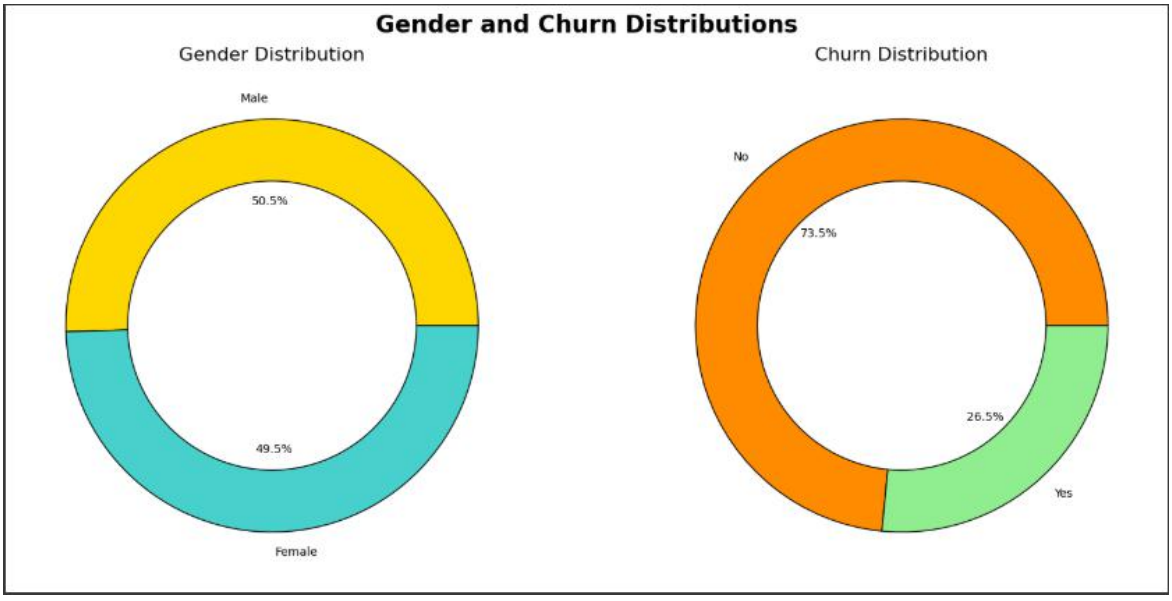
```
df.shape
```

```
(7043, 21)
```

```
df.info()
```

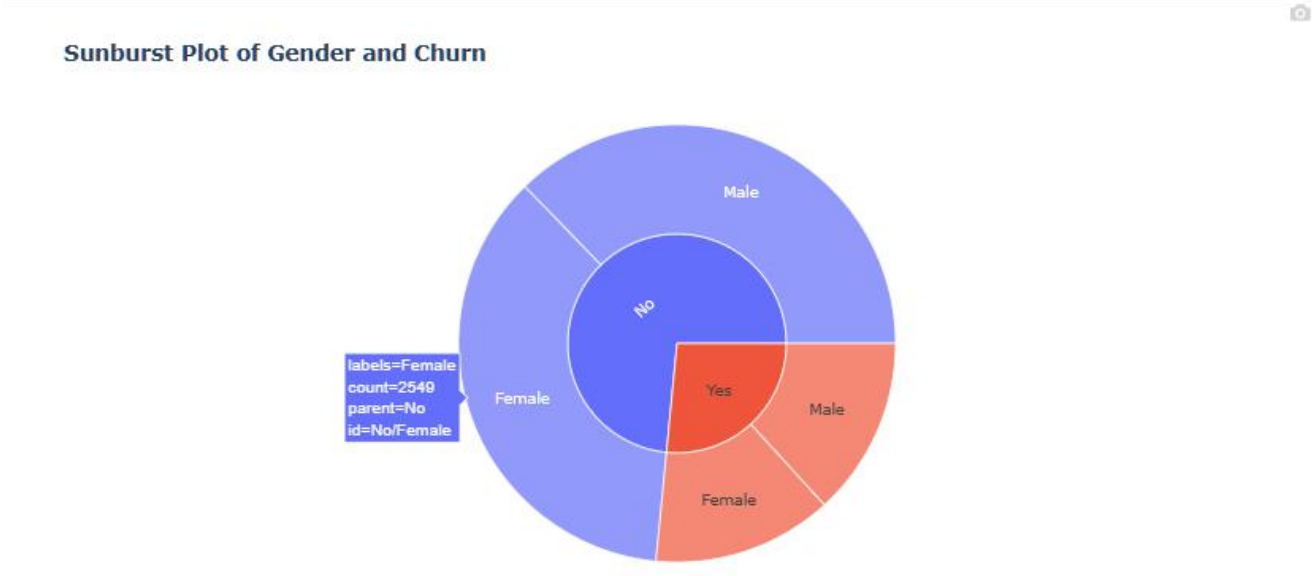
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7043 entries, 0 to 7042
Data columns (total 21 columns):
 #   Column            Non-Null Count  Dtype  
---  ------            --------------  -----  
 0   customerID        7043 non-null   object 
 1   gender            7043 non-null   object 
 2   SeniorCitizen     7043 non-null   int64  
 3   Partner           7043 non-null   object 
 4   Dependents        7043 non-null   object 
 5   tenure            7043 non-null   int64  
 6   PhoneService      7043 non-null   object 
 7   MultipleLines     7043 non-null   object 
 8   InternetService   7043 non-null   object 
 9   OnlineSecurity    7043 non-null   object 
 10  OnlineBackup      7043 non-null   object 
 11  DeviceProtection  7043 non-null   object 
 12  TechSupport       7043 non-null   object 
 13  StreamingTV       7043 non-null   object 
 14  StreamingMovies   7043 non-null   object 
 15  Contract          7043 non-null   object 
 16  PaperlessBilling  7043 non-null   object 
 17  PaymentMethod     7043 non-null   object 
 18  MonthlyCharges    7043 non-null   float64
 19  TotalCharges      7043 non-null   object 
 20  Churn             7043 non-null   object 
dtypes: float64(1), int64(2), object(18)
memory usage: 1.1+ MB
```

Gender and Churn Distributions

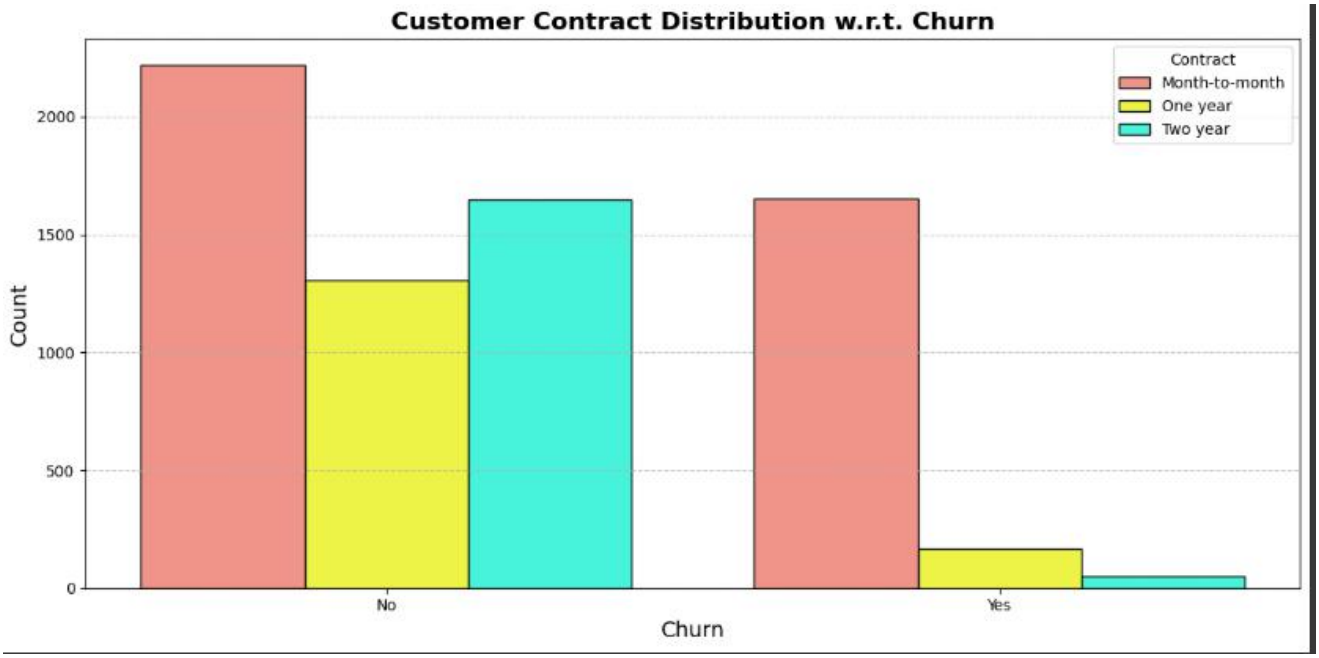


1. Customers are 49.5% Female and 50.5% Male.
2. 73.5% of them Churn whereas 26.5% of them don't.
3. Data is completely Imbalanced

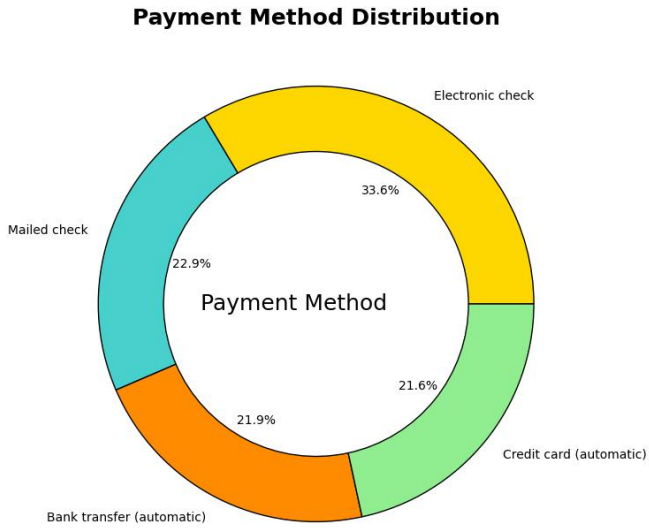
Sunburst plot of Gender and Churn



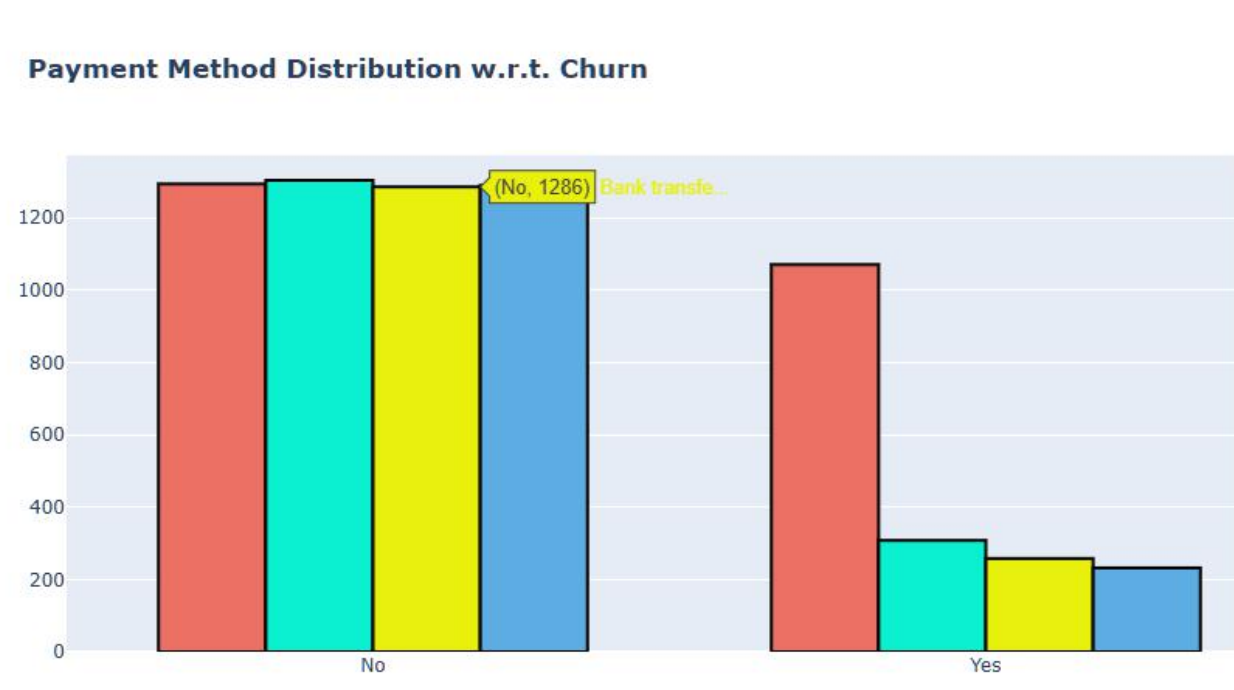
Customer Contract Distribution w.r.t Churn



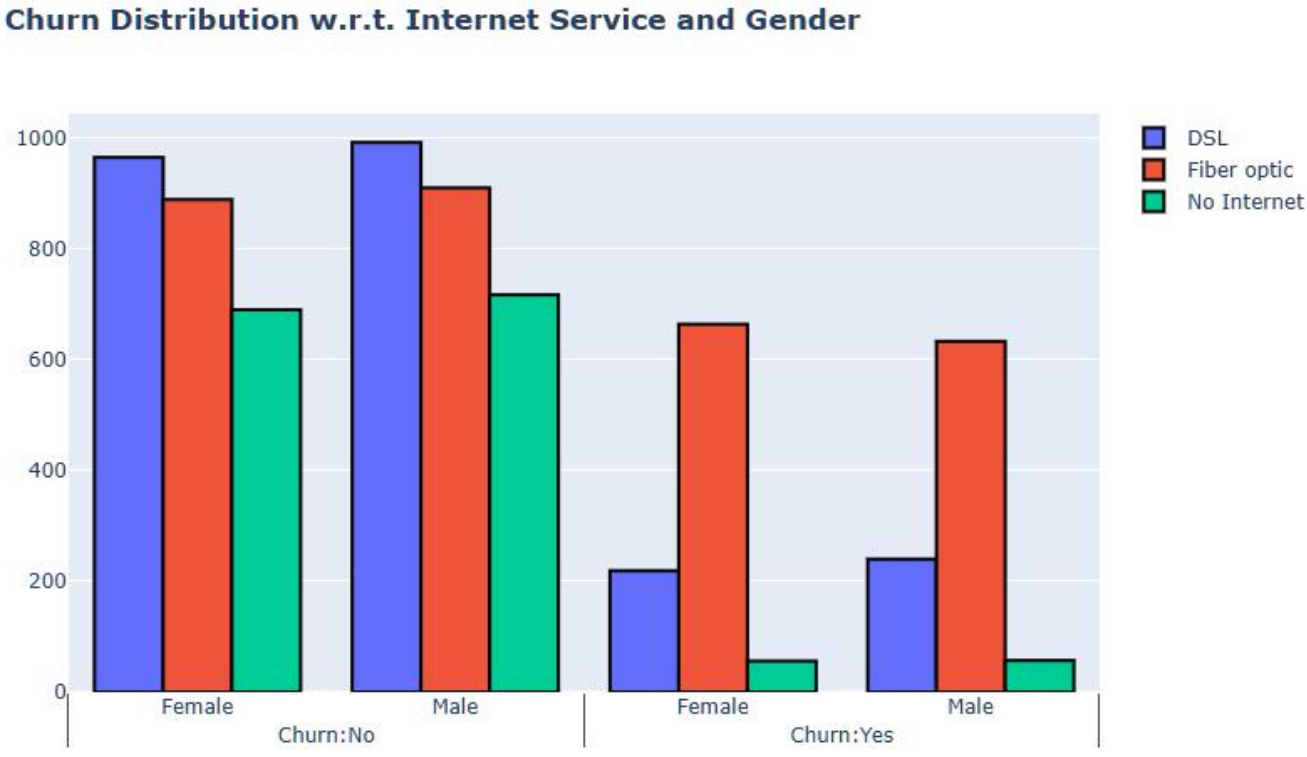
Payment Method Distribution



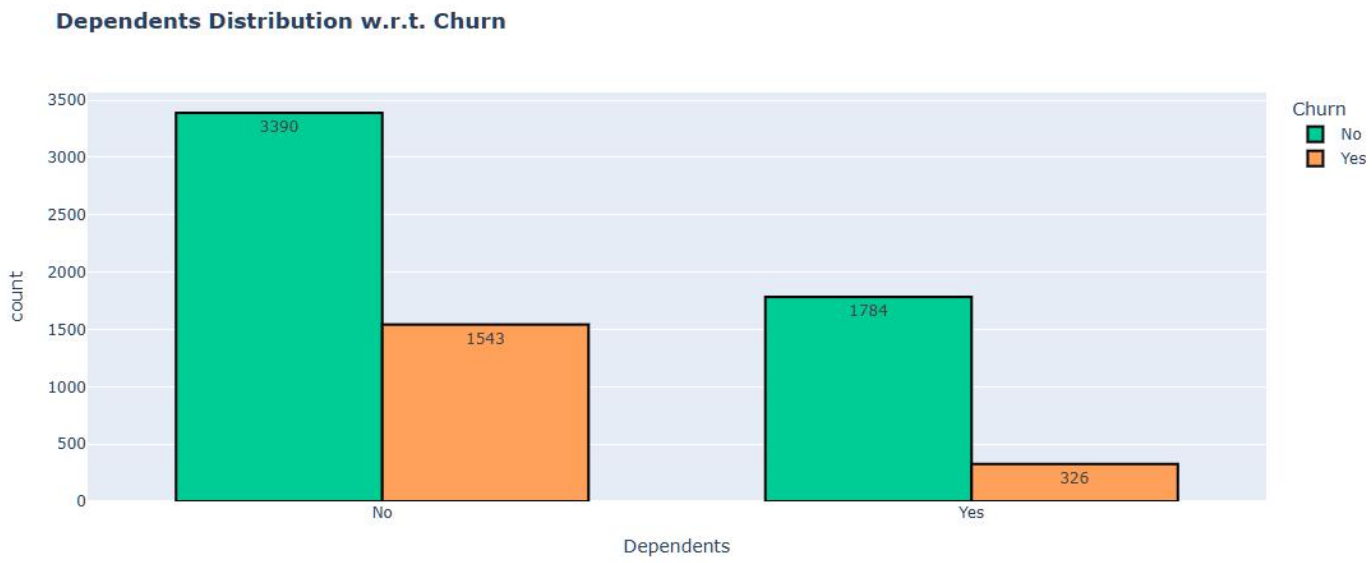
Payment Method Distribution w.r.t Churn



Churn Distribution w.r.t Internet Service and Gender



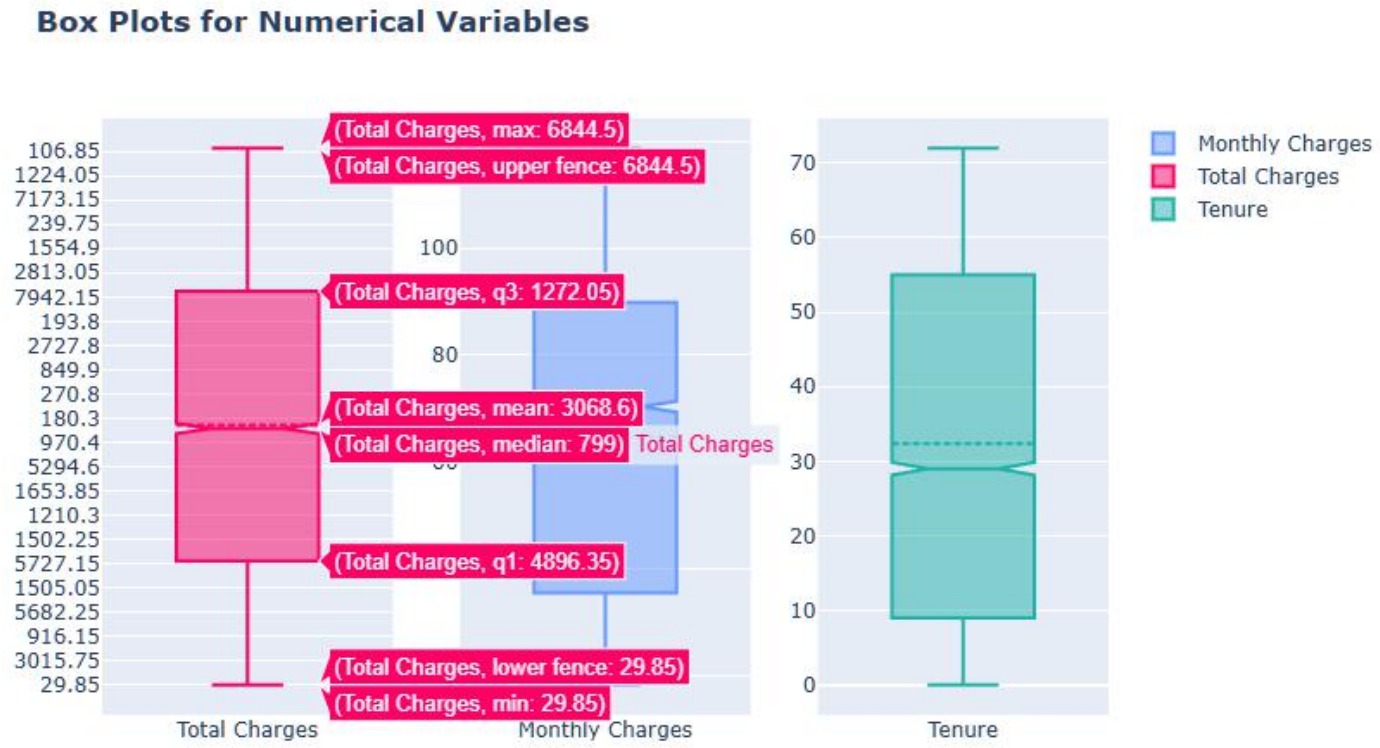
Dependents Distribution w.r.t Churn



Key takeways from Exploratory Data Analysis:

4. A female customer has a probability of 26.92 % churn
5. A male customer has a probability of 26.16 % churn
6. A customer with month-to-month contract has a probability of 42.71 % churn
7. A customer with one year contract has a probability of 11.27 % churn
8. A customer with two year contract has a probability of 2.83 % churn
9. A customer that uses Electronic check for paying has a probability of 45.29 % churn
10. A customer that uses Mailed check for paying has a probability of 19.11 % churn
11. A customer that uses Bank transfer (automatic) for paying has a probability of 16.71 % churn
12. A customer that uses Credit card (automatic) for paying has a probability of 15.24 % churn
13. A customer with dependents has a probability of 15.45 % churn
14. A customer without dependents has a probability of 31.28 % churn
15. A customer without a partner has a probability of 32.96 % churn
16. A customer with a partner has a probability of 19.66 % churn
17. A customer that is a senior citizen has a probability of 41.68 % churn
18. A customer that is not a senior citizen has a probability of 23.61 % churn
19. A customer with an online security has a probability of 14.61 % churn
20. A customer without an online security has a probability of 41.77 % churn
21. A customer with no internet service has a probability of 7.4 % churn
22. A customer with PaperlessBilling has a probability of 33.57 % churn
23. A customer without PaperlessBilling has a probability of 16.33 % churn
24. A customer with a tech support has a probability of 15.17 % churn
25. A customer without a tech support has a probability of 41.64 % churn
26. A customer with no internet service has a probability of 7.4 % churn
27. A customer with phone service has a probability of 26.71 % churn
28. A customer without phone service has a probability of 24.93 % churn

6. Box plots for Numerical Values



7. Categorical Encoding

```
df.head(5)
```

	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	OnlineB
0	1	0	0	1	1	0	0	1	2	
1	0	0	1	1	34	1	1	1	1	
2	0	0	1	1	2	1	1	1	1	
3	0	0	1	1	45	0	0	1	1	
4	1	0	1	1	2	1	1	2	2	

8. Splitting the dataset

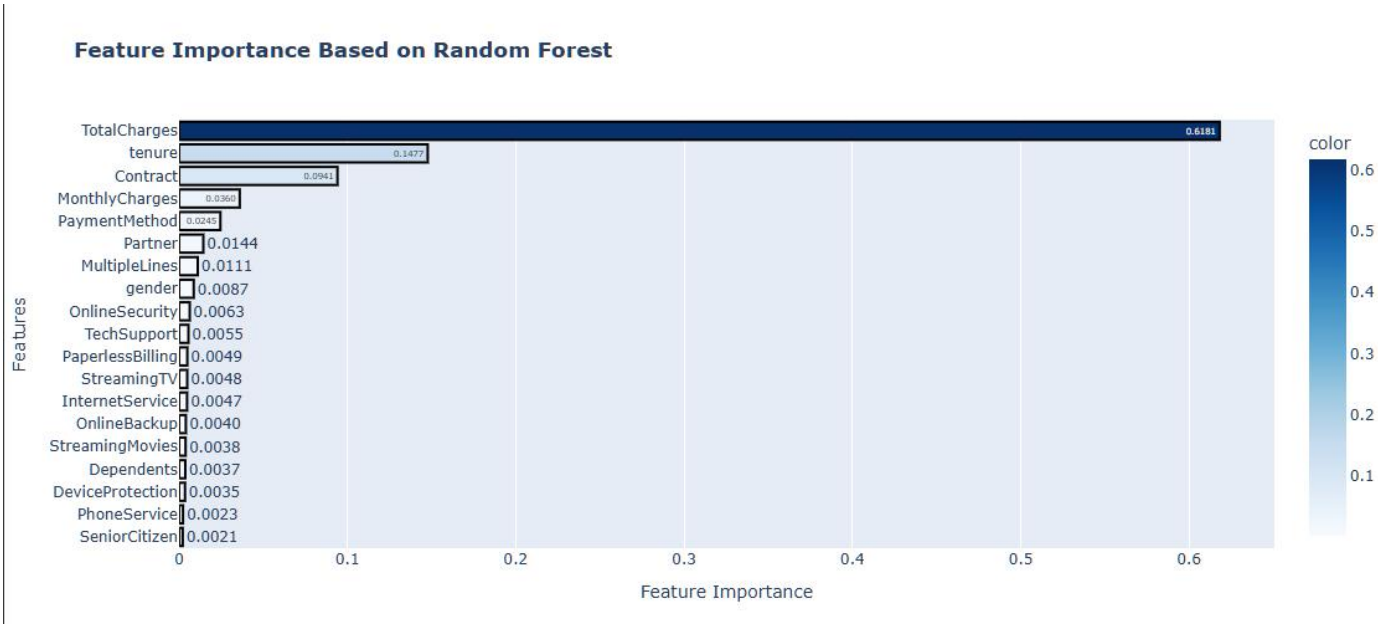
```
[50] X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.1, random_state=42)
     X_train.shape, X_test.shape

     ((9313, 19), (1035, 19))

[51] scaler = StandardScaler()
     X_train[['TotalCharges','MonthlyCharges','tenure']] = scaler.fit_transform(X_train[['TotalCharges','MonthlyCharges','tenure']])
     X_test[['TotalCharges','MonthlyCharges','tenure']] = scaler.transform(X_test[['TotalCharges','MonthlyCharges','tenure']])

[52] CV = StratifiedKFold(n_splits=10, random_state=0, shuffle=True)
```

9. Model Selection and Parameter Optimisation

Following machine learning models have been used and trained on

1. **Logistic Regression**
2. **K Nearest Neighbour**
3. **Random Forest**
4. **Decision Tree**
5. **AdaBoost**

10. Feature Importance



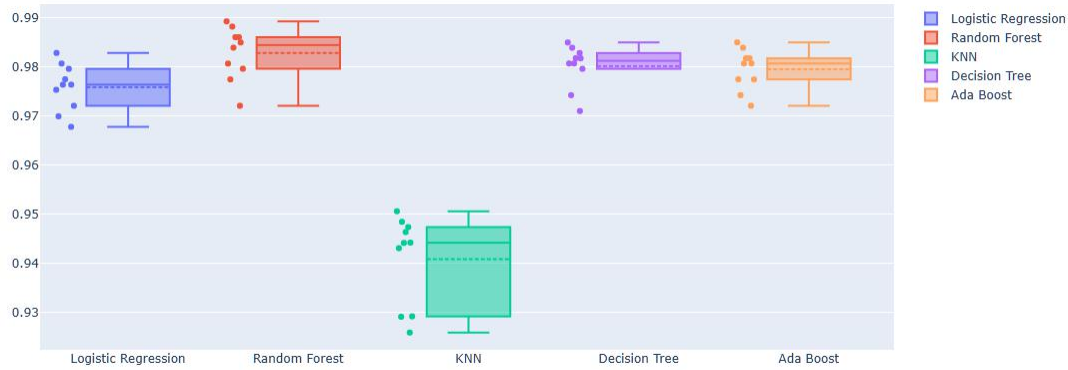
	Features	Feature Importance
1	SeniorCitizen	0.002121
5	PhoneService	0.002297
10	DeviceProtection	0.003482
3	Dependents	0.003655
13	StreamingMovies	0.003831
9	OnlineBackup	0.003989
7	InternetService	0.004710
12	StreamingTV	0.004803
15	PaperlessBilling	0.004856
11	TechSupport	0.005532
8	OnlineSecurity	0.006292
0	gender	0.008662
6	MultipleLines	0.011052
2	Partner	0.014375
16	PaymentMethod	0.024509
17	MonthlyCharges	0.035996
14	Contract	0.094059
4	tenure	0.147673
18	TotalCharges	0.618105

11. Results

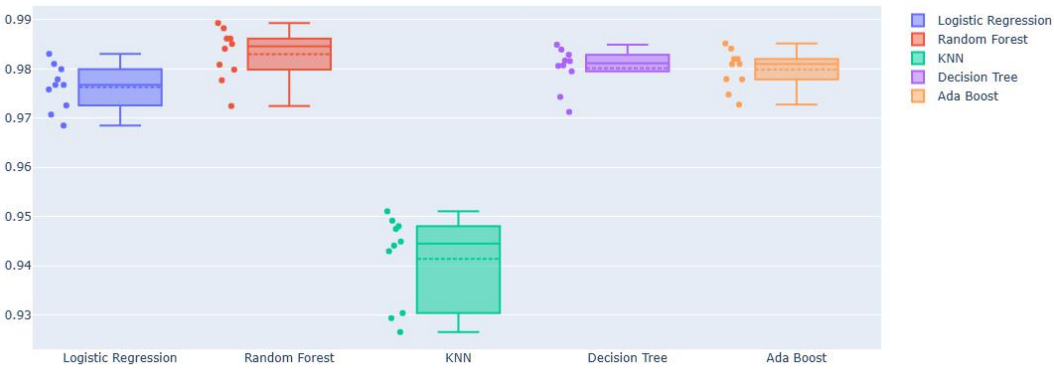
	Model	Accuracy Mean	F1 Score Mean	AUC Score Mean	Description
0	Logistic Regression	0.975839	0.976295	0.990839	
1	Random Forest	0.982819	0.982990	0.995353	
2	KNN	0.940836	0.941392	0.940867	
3	Decision Tree	0.980135	0.980137	0.981270	
4	Ada Boost	0.979491	0.979863	0.991575	

\

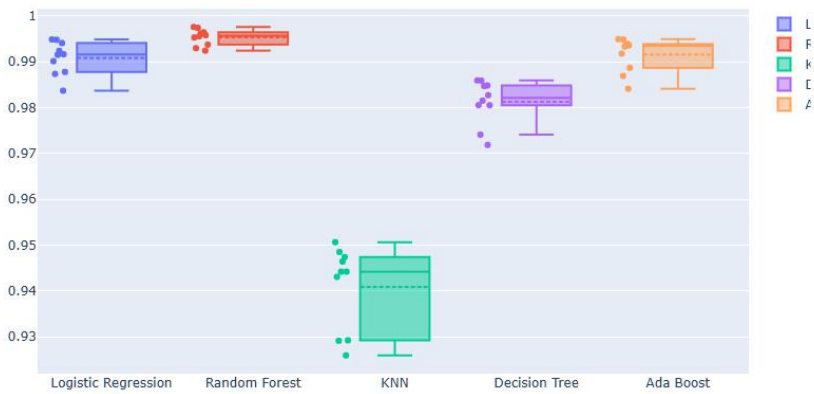
Box plots for Model Accuracy (train)



Box plots for Models F1 score (train)



Box plots for Models AUC (train)



Classification Report

```
RF_model = RandomForestClassifier(random_state=42)
RF_model.fit(X_train, y_train)
y_pred = RF_model.predict(X_test)
```

```
[76] print(classification_report(y_test,y_pred))
```

```
              precision    recall  f1-score   support

           0       1.00      0.97      0.98       505
           1       0.97      1.00      0.99       530

    accuracy                           0.99      1035
   macro avg       0.99      0.99      0.99      1035
weighted avg       0.99      0.99      0.99      1035
```

6. Applicable Regulations and Limitations

In developing and offering our SaaS application, it is essential to adhere to various regulations that govern data usage, privacy, and compliance. Key regulations include the General Data Protection Regulation (GDPR) for users in the European Union, the California Consumer Privacy Act (CCPA) for users in California, and other local data protection laws. These regulations mandate that we obtain explicit consent from users before collecting, processing, or storing their personal data. While our SaaS application offers machine learning capabilities, there are limitations to consider:

6.1 **Dependence on Third-Party APIs:** If our application integrates with the APIs of other products, we may face limitations based on those APIs' availability, performance, and terms of use. Any changes made by these third-party providers can directly affect our service delivery and user experience.

6.2 Data Integrity and Quality: The effectiveness of our machine learning models relies heavily on the quality and integrity of the data provided by users. Inaccurate, incomplete, or outdated data can lead to misleading insights and recommendations, potentially impacting business decisions.

6.3 Compliance with Data Transfer Regulations: When dealing with cross-border data transfers, we must ensure compliance with applicable regulations regarding data protection. This may involve implementing standard contractual clauses or other mechanisms to safeguard user data during transfer.

6.4 Model Limitations: While our machine learning models can provide valuable insights, they are not infallible. Users should be aware that models may not account for all variables or changes in business environments, and human judgment should always complement automated recommendations.

6.5 User Education and Awareness: It's vital for us to educate our users about the importance of data privacy and security. By ensuring that users understand their responsibilities and the best practices for data management, we can help them maintain compliance and protect their customers' information.

6.6 Data Retention Policies: We must establish clear data retention policies to comply with regulations and ensure that user data is only kept as long as necessary

7. Business Model

7.1 Subscription-Based Pricing:

7.1.1 Tiered Plans: Offer multiple subscription tiers (e.g., Basic, Professional, and Enterprise) with varying levels of access to features, data storage, and support. This allows businesses of different sizes and needs to choose a plan that fits their requirements.

7.1.2 Monthly or Annual Billing: Provide options for monthly or annual billing, with discounts for annual subscriptions to encourage long-term commitments.

7.2 Cost of API Access:

7.2.1 API Access Fees: Implement a structured pricing model for API usage, where users pay based on the number of API calls made or the volume of data retrieved. This could include different tiers that cater to varying usage levels, such as a limited free tier for testing and development, followed by graduated pricing for higher usage. This approach allows businesses to integrate your machine learning services seamlessly into their systems while giving them control over their costs based on their specific needs.

7.3 Usage-Based Pricing:

7.3.1 Pay-as-You-Go Model: Charge users based on their actual usage of the application, such as the number of API calls made or the volume of data processed. This model is attractive to businesses that want to scale their usage without committing to a fixed price.

7.4 Freemium Model:

7.4.1 Basic Free Tier: Offer a basic version of the application for free, allowing users to access limited features. This strategy helps attract a larger user base and provides an opportunity to upsell premium features or plans.

7.4.2 In-App Purchases: Provide additional features, advanced analytics, or personalized support as add-ons that users can purchase within the application.

7.5 Consulting and Support Services:

7.5.1 Custom Implementation Services: Offer consulting services to help businesses integrate the application with their existing systems or customize it to meet their specific needs.

7.5.2 Training and Support Packages: Provide paid training sessions, workshops, or dedicated support for users who want assistance in maximizing the value of the application.

7.6 Affiliate and Partnership Programs:

7.6.1 Referral Programs: Create a referral program that rewards existing users for bringing in new customers, creating a win-win scenario for both parties. Collaborate with complementary software providers to offer bundled services or integrations, potentially generating additional revenue through joint marketing efforts

8 Final Product Prototype Details

The ultimate product prototype of our Customer Insights Analysis SaaS Application is designed to revolutionize how businesses leverage data for decision-making. This platform seamlessly integrates advanced machine learning algorithms to provide actionable insights on customer behavior, churn prediction, and sentiment analysis. By fostering community engagement and collaborating with eCommerce and retail businesses, this application empowers organizations to enhance customer satisfaction and drive growth. This overview is complemented by a detailed schematic diagram illustrating the system architecture and key functionalities.

8.1 Abstract

Our Customer Insights Analysis SaaS Application is a comprehensive platform designed to empower businesses to harness the power of data for improved decision-making and customer engagement. Utilizing advanced machine learning algorithms, this application delivers valuable insights into customer behavior, churn prediction, and sentiment analysis, enabling businesses to tailor their strategies for maximum impact.

From a technical perspective, our application is built using a modern technology stack, including React for the frontend to ensure a responsive and intuitive user interface, and Flask for the backend to manage data processing and API integrations efficiently. Leveraging machine learning frameworks like TensorFlow and scikit-learn, we empower businesses with predictive analytics capabilities. . Our product is offered as a SaaS application, providing businesses with a user-friendly platform to access analytics and insights, while also offering external API access for seamless integration into existing workflows and systems.

A key advantage of our application is its integration capabilities with various eCommerce platforms and POS systems, allowing businesses to seamlessly access and analyze their data. By partnering with local businesses, we provide a network of resources that enhances visibility and facilitates better decision-making. This collaborative approach not only strengthens community ties but also enables businesses to leverage shared insights and best practices for continuous improvement. Moreover, our application supports customizable dashboards and reporting tools that provide actionable metrics tailored to each business's specific needs, ensuring that users can make informed decisions quickly and effectively.

Conclusion

In summary, the Customer Insights Analysis SaaS application offers a practical solution for businesses looking to better understand their customers. By combining features like customer churn prediction, segmentation, sentiment analysis, and review scoring, we provide valuable insights that help businesses make smarter decisions and improve customer satisfaction. The user-friendly design ensures that businesses can easily navigate the platform and access the information they need. With real-time data updates and seamless API integration, our application helps users stay responsive to their customers' needs. By focusing on the specific requirements of eCommerce sites, POS systems, and marketplace sellers, we aim to deliver relevant insights that can drive meaningful change. Overall, the Customer Insights Analysis SaaS application is about empowering businesses with the knowledge they need to enhance their customer relationships and achieve better outcomes

Repo Link: https://github.com/sharaaan/Customer-Churn-Rate-Project-3-Sharan-P.git