**PHASE 5**

**PROJECT DOCUMENTATION & SUBMISSION**

|  |  |
| --- | --- |
| **Date** | **31-10-2023** |
| **Team ID** | **721** |
| **Project Name** | **Customer Churn Prediction** |

**Project Title:** Customer Churn Prediction

**Table of Contents**

|  |  |
| --- | --- |
| 1 | Introduction |
| 2 | Problem Statement |
| 3 | Problem identified |
| 4 | Key Challenges |
| 5 | Literature Survey |
| 6 | Design Thinking |
| 7 | Technology Architecture |
| 8 | Steps in Model Evaluation |
| 9 | Predictive Model |
| 10 | Conclusion |

**1.INTRODUCTION**

Customer churn, or the loss of customers, poses a significant challenge for businesses operating in diverse industries. What makes this issue even more pressing is the widely recognized fact that the cost of acquiring a new customer is substantially higher—often estimated to be 5-25 times more—than the cost of retaining an existing one. This substantial cost disparity underscores the compelling need for businesses to identify and preserve their most valuable customers.

One effective approach to mitigating customer churn involves harnessing the power of predictive analytics. This is achieved by developing a machine learning model that leverages historical customer data, complete with information regarding whether customers have churned or not. This historical data forms the foundation for training a predictive model. Once this model is trained and validated, it is primed for use in forecasting the churn risk for individual customers. Armed with these predictions, businesses can proactively craft and implement retention strategies aimed at preserving their customer base. These strategies might include tailored incentives like discounts or personalized outreach to re-engage and satisfy customers at risk of churning.

The project seeks to build a robust machine learning model with the capability to foresee customer churn. The model will be trained on a comprehensive dataset of historical customer information gathered from various industries. This dataset encompasses a wide range of attributes and behaviours to comprehensively understand customer dynamics.

Once the model is developed, it will undergo rigorous evaluation using a variety of performance metrics such as accuracy, precision, and recall. These metrics serve to gauge the model's effectiveness in predicting churn. Concurrently, the project also entails a thorough analysis of the key factors that influence customer churn. Feature engineering and data analysis techniques will be applied to uncover the underlying factors driving customer attrition.

Upon successful development and validation of the model, the next crucial step is its deployment to a production environment. This operationalizes the model, making it readily accessible for businesses to employ in their efforts to identify and retain their most valuable customers. With the model in place, businesses can make data-driven decisions and execute proactive measures to maintain customer loyalty and minimize churn.

In summary, this project constitutes a pivotal endeavour for businesses seeking to enhance customer retention and reduce churn. It harnesses predictive analytics and machine learning to not only predict customer churn but also to unearth the factors that underlie it. The end goal is to empower businesses to take informed actions that preserve their most valuable customers and, in turn, optimize their bottom line.

**2.PROBLEM STATEMENT:**

Objective: The primary objective of this project is to address the pervasive issue of customer churn within businesses. Customer churn, or the loss of customers, presents a significant challenge that can lead to revenue loss and increased customer acquisition costs. To mitigate this problem, the project aims to develop a robust machine learning model. This model's core function is to accurately predict which customers are at risk of churning. It achieves this by leveraging historical customer data, including whether customers have previously churned or not. Furthermore, the project seeks to delve deeper into the factors that exert the most significant influence on customer retention. Through feature engineering and comprehensive data analysis, it aspires to uncover and understand the key drivers that impact whether customers remain loyal or decide to leave. By achieving these objectives, the project equips businesses with the insights and tools they need to proactively manage churn, retain valuable customers, and ultimately enhance their bottom line.

**3.PROBLEM IDENTIFIED:**

**Customer churn is a major cost to businesses.** **Predicting churn and identifying key retention factors is critical.** **This project aims to build a predictive model using machine learning.** **The model will be trained on historical customer data and churn labels.** **The key retention factors will be identified through feature engineering and analysis.** **The model will be used to predict the churn risk of individual customers.** **Businesses can use these predictions to proactively implement retention strategies.**

**4.KEY CHALLENGES:**

1. Data Quality: Ensuring the dataset is clean, complete, and free of errors.

2. Feature Selection: Identifying the most relevant customer attributes and usage patterns for accurate churn prediction.

3. Model Selection: Choosing appropriate data analytics techniques and algorithms for the task.

4. Model Evaluation: Evaluating the model's performance using appropriate metrics.

5. Recommendations: Providing actionable insights to businesses for reducing customer attrition.

**5.LITERATURE SURVEY**

1. **“Customer Churn Prediction in the Banking Sector Using Machine Learning Based Classification Models”, Hoang Tran, Ngoc Le, Van-Ho Nguyen [2022]**

The initial goal was to examine the impact of customer segmentation on the accuracy of customer churn prediction in the banking sector using machine learning models. The second objective is to experiment, contrast, and assess which ma-chine learning approaches are most effective in predicting customer churn. This paper reviews the theoretical basis of customer churn, and customer seg-mentation, and suggests using supervised machine-learning techniques for customer attrition prediction. In this study, we use different machine learning models such as k-means clustering to segment customers, k-nearest neighbours, logistic regression, decision tree, random forest, and support vector machine to apply to the dataset to predict customer churn.

1. **“Customer churn prediction for a webcast platform via a voting-based ensemble learning model with Nelder-Mead optimizer”, Kani Fu, Guiyang Zheng, Wei Xie [2023]**

This article studies the application for customer churn prediction on webcast. Predicting churn customers become an urgent need in webcast industry because the market is getting saturated and identifying potential churn customers and developing recall marketing strategies can save companies significant costs. Despite the importance of customer churn prediction in many fields, little prior academic attention has been attached to the webcast area. To address this gap, we apply an ensemble learning method to build a binary classification model for customer churn prediction. Our proposed model uses a weighted voting ensemble method and the Nelder-Mead optimal algorithm with a specific focus on the speed of Internet customers’ mobility, extracting high-dimensional features from time series data to incorporate more detailed customer behaviour information. In addition, a new customer churn indicator based on time decline is introduced to more accurately define churned customers in the training data. The experimental data is collected from a webcast application developed by a Chinese Internet company. Experimental evaluations show that compared to the traditional ensemble models, our proposed model is operationally efficient and outperforms other approaches, providing valuable insights for companies to intervene with churned customers and adopt targeting retention interventions.

**3.“Data Visualization and Prediction for Telecom Customer Churn”, Pulin Yang [2023]**

With the deepening of telecom industry reform and the intensification of competition, the customer churn rate of telecom enterprises is gradually increasing. How to predict and effectively reduce customer churn is directly related to the survival and development of telecom enterprises. In order to effectively deal with unbalanced classification and improve the accuracy of high-value customer churn prediction in telecom industry, this paper uses telecom customer data set from Kaggle platform to analyse people's use of telecom services, and help telecom operators find out the reasons for customer churn, and establish churn prediction model to reduce customer churn rate. In this paper, firstly, the data set is imported, and then the data visualization analysis is carried out. Then, the random forest model, SVM model and GBDT model are introduced for comparison. Experiments show that random forest has better classification performance than other methods, and improves the accuracy of high-value customer churn prediction.

**4.“Prediction of Customer Churn on e-Retailing”, M Jaeyalakshmi, S Gnanavel, K S Guhapriya, S Harshini Phriyaa , K Kavya Sree [2020]**

The technology has always been an instigating factor in progress for human civilization which resulted in driving the customer services to a greater need. The enrichment of technology has amplified and embellished the customer interaction among various business to consumer sectors. These technological upgrading have a huge impact on the retail industry which is an ever-growing market with key competitors around the world. In a consortium of multiple competitors in the same business, the re-engagement of disinterested customers is essential rather than winning a new customer. The sustenance of a customer can be figure out by Churn Prediction. Churn prediction is a new promising method in customer relationship management to analyse customer retention in subscription-based business. It is the activity of identifying customer with a high probability to discontinue the company based on analysing their past data and behaviour. It looks at what kind of customer data are typically used, do some analysis of the features chosen, and initiate a churn prediction model. Thus, churn prediction is a valuable approach in identifying and profiling the customers at risk.

**5.“Machine Learning to Develop Credit Card Customer Churn Prediction”, Dana AL-Najjar, Nadia Al-Rousan, Hazem AL-Najjar [2022]**

The credit card customer churn rate is the percentage of a bank’s customers that stop using that bank’s services. Hence, developing a prediction model to predict the expected status for the customers will generate an early alert for banks to change the service for that customer or to offer them new services. This paper aims to develop credit card customer churn prediction by using a feature-selection method and five machine learning models. To select the independent variables, three models were used, including selection of all independent variables, two-step clustering and k-nearest neighbour, and feature selection. In addition, five machine learning prediction models were selected, including the Bayesian network, the C5 tree, the chi-square automatic interaction detection (CHAID) tree, the classification and regression (CR) tree, and a neural network. The analysis showed that all the machine learning models could predict the credit card customer churn model. In addition, the results showed that the C5 tree machine learning model performed the best in comparison with the three developed models. The results indicated that the top three variables needed in the development of the C5 tree customer churn prediction model were the total transaction count, the total revolving balance on the credit card, and the change in the transaction count. Finally, the results revealed that merging the multi-categorical variables into one variable improved the performance of the prediction models.

**6.DESIGN THINKING**

**Design Thinking Approach**

**Empathize:**

Before solving the problem, it's crucial to empathize with businesses and understand their pain points related to customer churn. We need to gather insights into the industries, customer segments, and reasons why customers leave.

**Actions:**

- Conduct interviews or surveys with business stakeholders to understand their challenges related to customer churn.

- Analyse historical churn data to identify common patterns and reasons for attrition.

- Seek input from customer support and sales teams for anecdotal insights.

**Define:**

Based on our understanding of the problem and business needs, we will define clear objectives and success criteria for our project.

**Objectives:**

- Develop a data analytics model that achieves a high accuracy rate in predicting customer churn.

- Identify the top factors influencing customer retention.

- Provide actionable recommendations to reduce churn based on the analysis.

**Ideate:**

Brainstorm potential solutions and approaches to address the problem. This phase involves creatively thinking about data analytics techniques and methods for churn prediction.

**Actions:**

- Explore various data analytics methods, including logistic regression, decision trees, random forests, and clustering.

- Experiment with feature engineering to create relevant customer attributes.

- Consider sentiment analysis on customer feedback and social media data to gauge customer satisfaction.

**Prototype:**

Create a prototype of the data analytics model and the analytics dashboard for churn analysis.

**Actions:**

- Develop data preprocessing scripts and analytics notebooks to build and evaluate predictive models.

- Create an interactive dashboard using tools like Tableau or Power BI to visualize churn-related insights.

- Test the prototype using a subset of the dataset to ensure it meets performance objectives.

**Test:**

Evaluate the model's performance using appropriate metrics and gather feedback from business stakeholders.

**Actions:**

- Split the dataset into training and testing sets.

- Train the data analytics model on the training set and evaluate it on the testing set.

- Use metrics such as accuracy, precision, recall, and F1-score to assess model performance.

- Collect feedback from business stakeholders on the dashboard's usability and insights.

**Implement:**

Once the prototype meets the defined objectives and receives positive feedback, proceed with full implementation.

**Actions:**

- Train the final data analytics model on the entire dataset.

- Deploy the model and analytics dashboard as part of a production-ready solution.

- Conduct thorough testing to ensure the system is robust and user-friendly.

**Iterate:**

Continuous improvement is essential. Gather business feedback and iterate on the model and dashboard to enhance accuracy and usability.

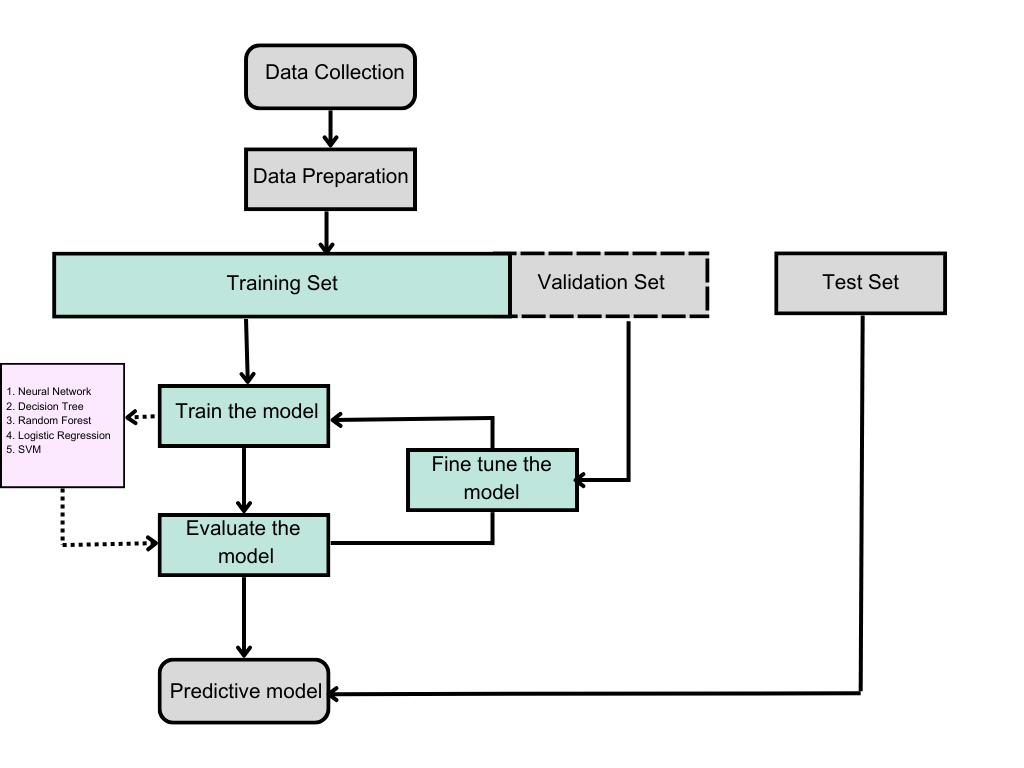
**Actions:**

- Monitor the model's performance and retrain it periodically with updated customer data.

- Address business feedback and make necessary improvements to the analytics dashboard.

- Stay informed about advancements in data analytics and customer churn prediction techniques for potential enhancements.

**7.TECHNOLOGY ARCHITECTURE:**



**7.1 Data Collection and Feature Engineering:**

Innovation: Comprehensive Data Gathering

Utilize various data sources, including customer demographics, transaction history, and customer interactions, to collect comprehensive datasets.

Apply advanced data feature engineering techniques, such as customer segmentation, to extract meaningful insights from the data.

Create new features like customer lifetime value, customer satisfaction scores, and interaction frequency to improve churn prediction.

**7.2 Data Pre-processing**

Innovation: Natural Language Processing (NLP) for Feedback Analysis

Employ NLP techniques to analyse customer feedback and reviews.

Develop a custom NLP pipeline for sentiment analysis, topic modelling, and identifying specific customer concerns.

Handle missing data and outliers using innovative methods, such as imputation based on customer profiles and anomaly detection.

**7.3 Model Selection and Training**

Innovation: Advanced Machine Learning Models

Utilize a range of machine Learning algorithms, including Logistic Regression, Decision trees, Random Forest, and Support Vector Machines, for churn prediction.

Explore ensemble methods, such as Bagging and Boosting, to combine multiple models for improved accuracy.

Implement Deep Learning techniques, like Neural Networks, to capture complex patterns in customer behaviour.

**7.4 Customer Segmentation**

Innovation: Behavioural Clustering

Apply Unsupervised Learning techniques to segment customers based on their behaviour and characteristics.

Develop innovative Clustering algorithms that consider temporal patterns and transactional data for better segmentation.

Tailor retention strategies for each customer segment to increase effectiveness.

**7.5 Customer Engagement Analysis**

Innovation: Interaction History Analysis

Analyse historical customer interactions, including emails, calls, and website visits, to understand engagement patterns.

Implement innovative techniques like sequence analysis to identify customer touchpoints and their impact on churn.

Use Reinforcement Learning to optimize customer engagement strategies.

**7.6 Explainable AI (XAI)**

Innovation: Model Interpretability

Employ XAI techniques, such as SHAP values and LIME, to provide interpretable explanations for churn predictions.

Create a user-friendly dashboard with visual explanations to enhance decision-making for retention strategies.

**7.7 Continuous Learning**

Innovation: Real-time Model Updates

Establish a Continuous Learning framework that adapts to changing customer behaviours and market conditions.

Regularly retrain the model using new data and customer feedback.

Implement automated data pipelines for seamless data ingestion and model updates.

**8.STEPS INVOLVED IN MODEL EVALUATION**

**Step 1:**

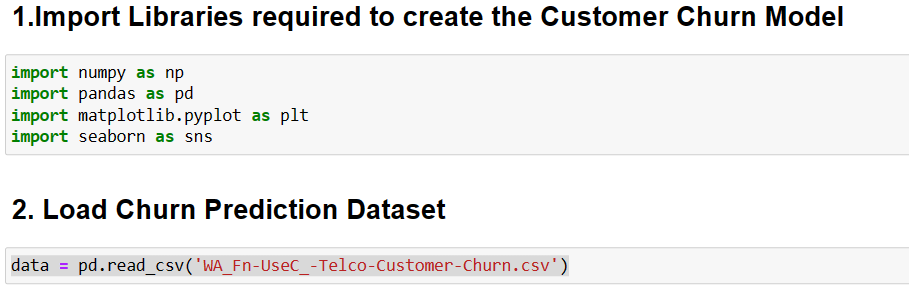
**Data collection:**

The data collection begins by defining the project's objectives and understanding the problem domain. Next, data sources are identified, which could range from databases and APIs to surveys or sensors. The actual gathering process involves extracting or obtaining the data from these identified sources. This stage involves defining project objectives, identifying relevant data sources, gathering data, preprocessing (handling missing values, outliers, etc.), cleaning, validating, exploring, and visualizing the data.

**Step 2:**

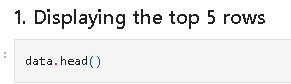
**Load the dataset.**

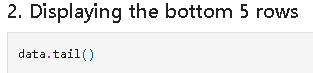
Import the necessary dependencies and load the dataset.

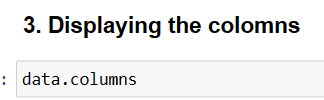


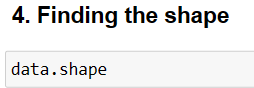
**Step 3:**

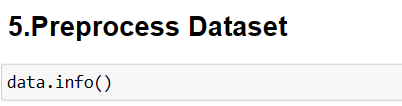
**Explore the dataset using jupyter notebook.**

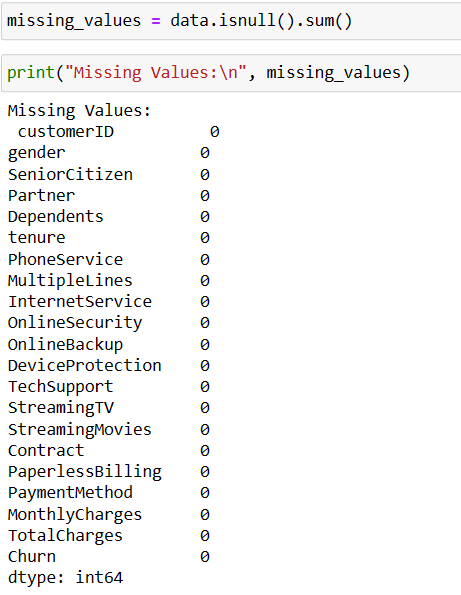
****

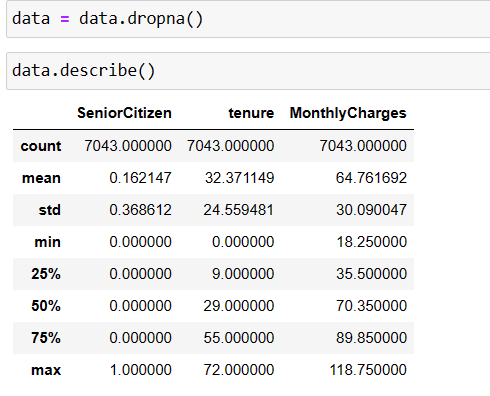
****

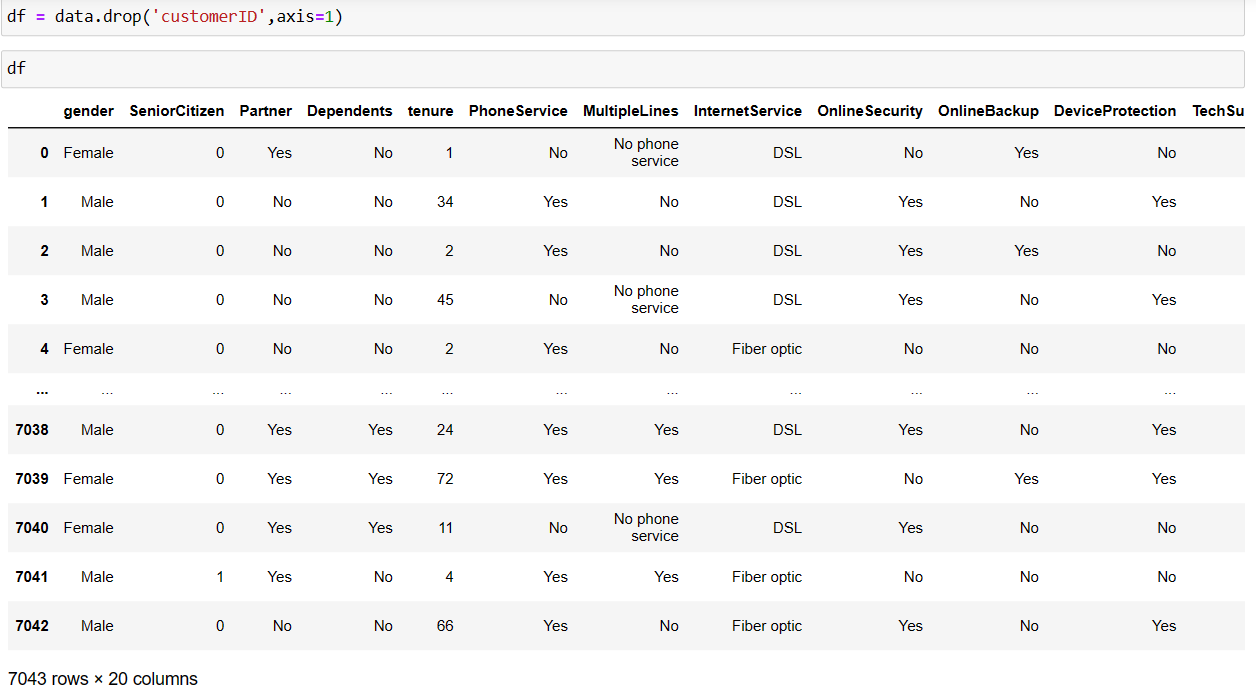
****

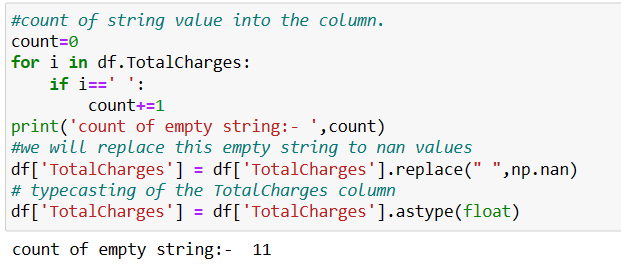
****

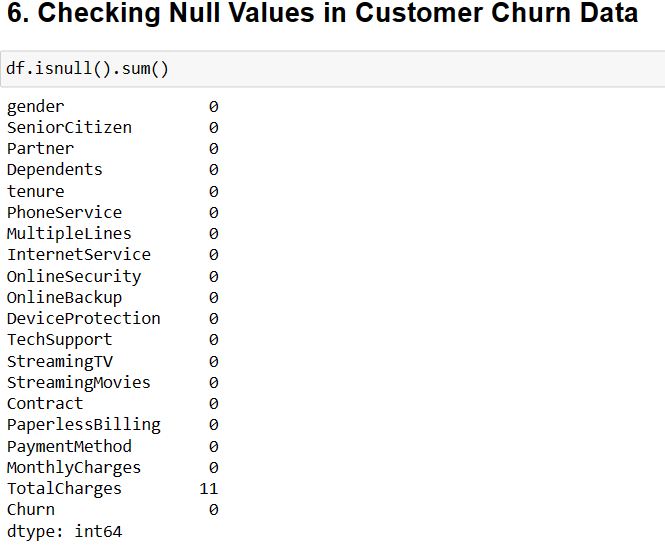
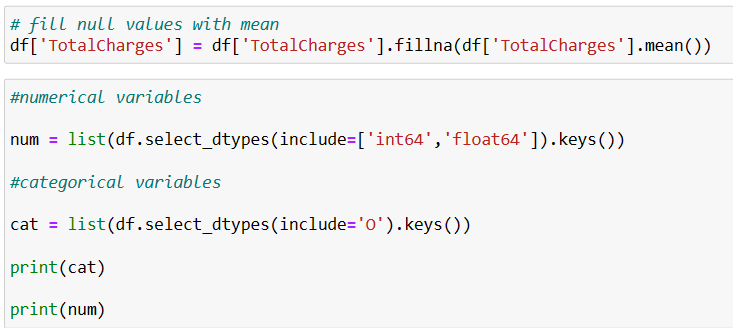
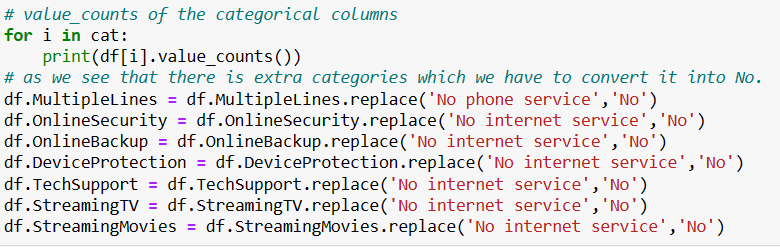
****

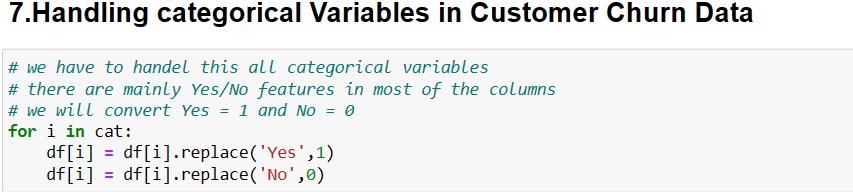
****

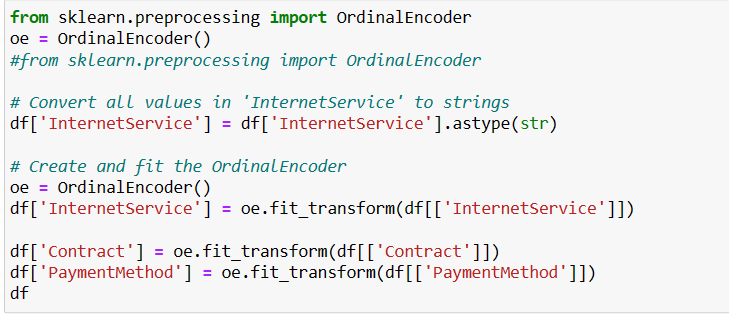
****

****

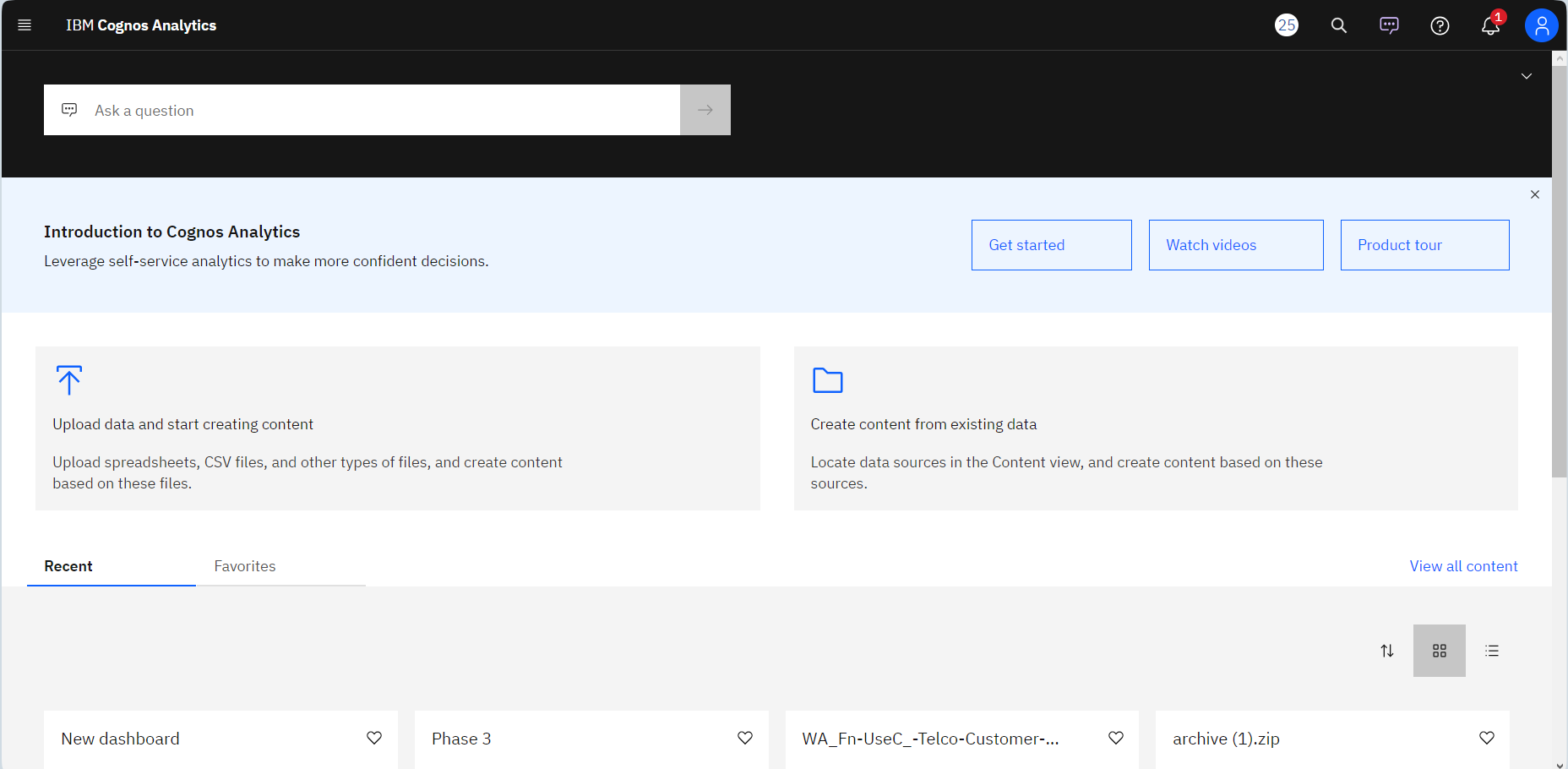
****

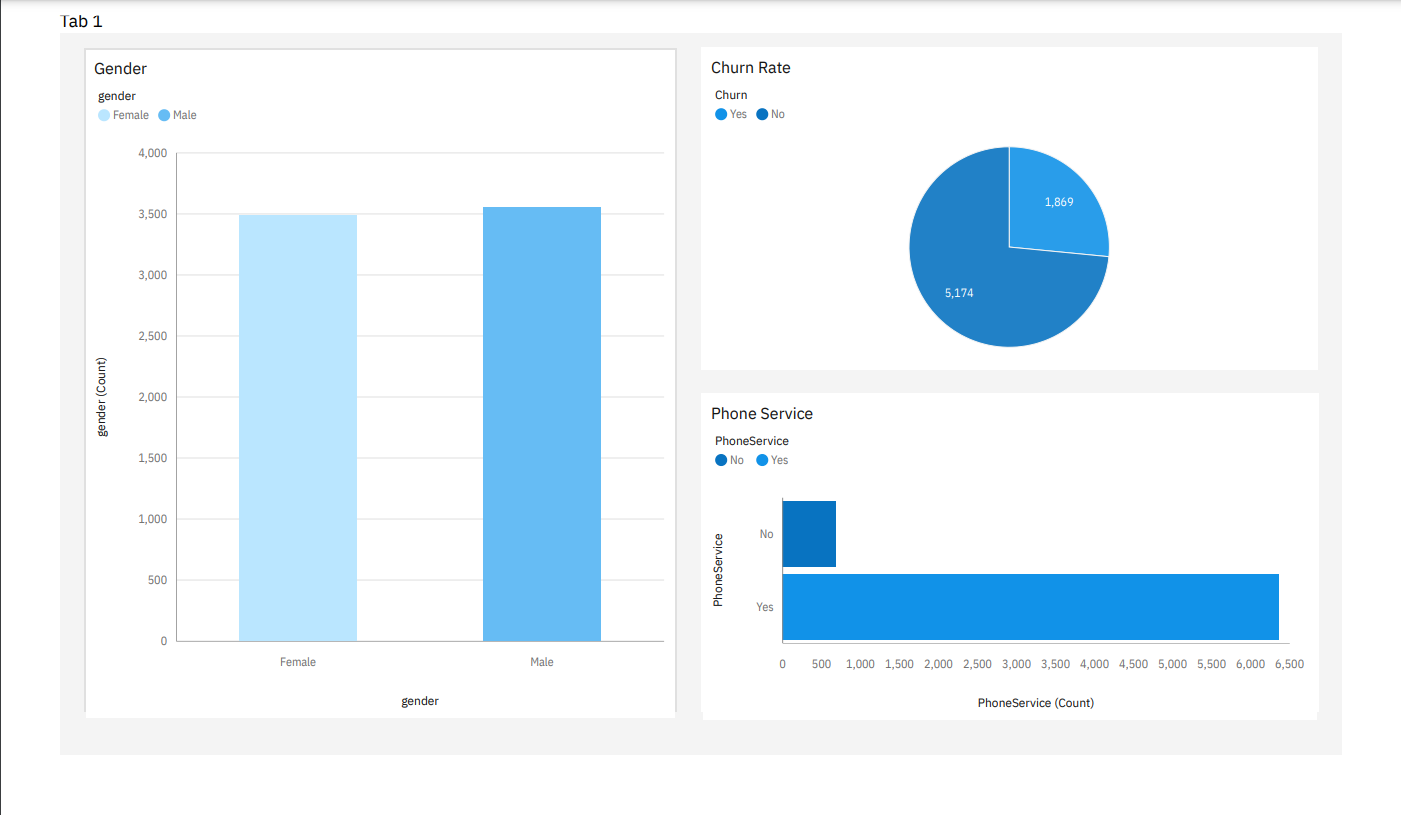


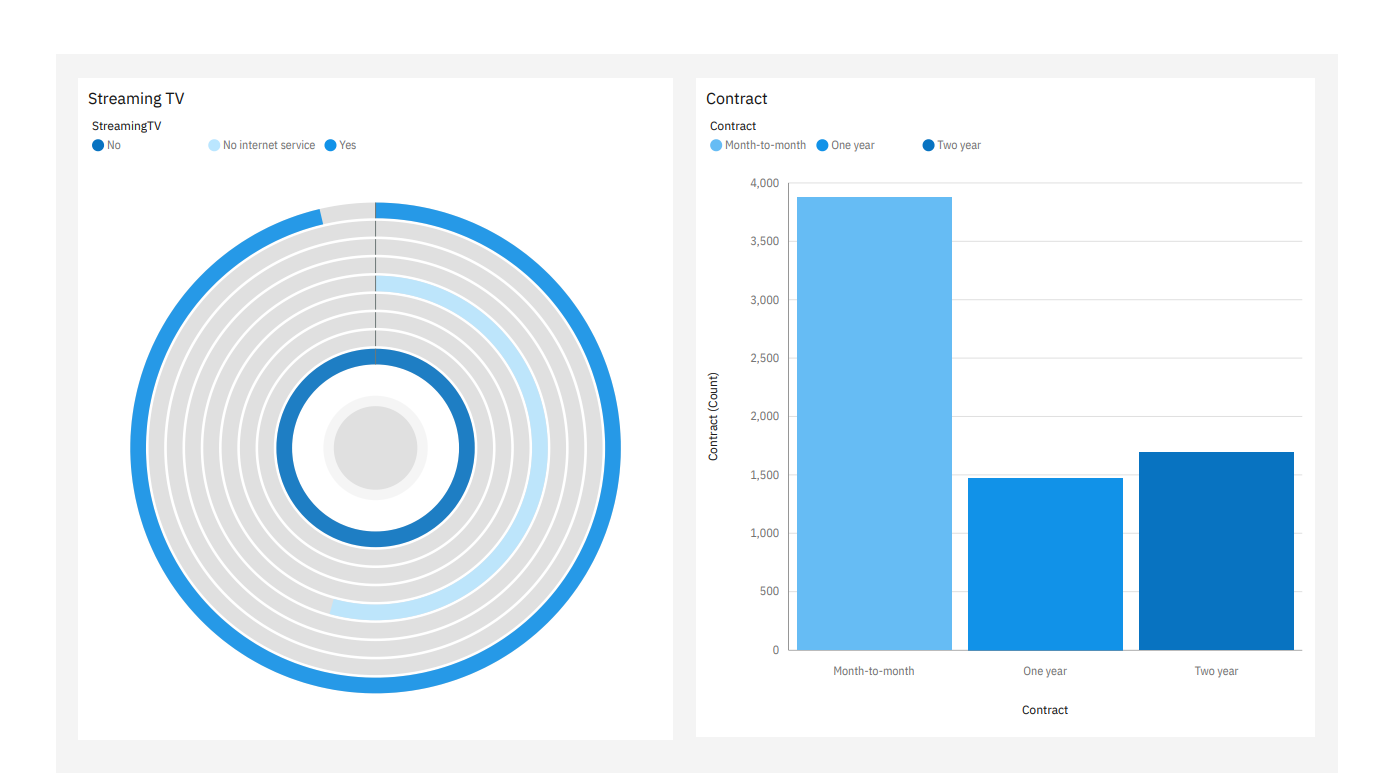


**8. Account creation in IBM Cognos Analytics**

****

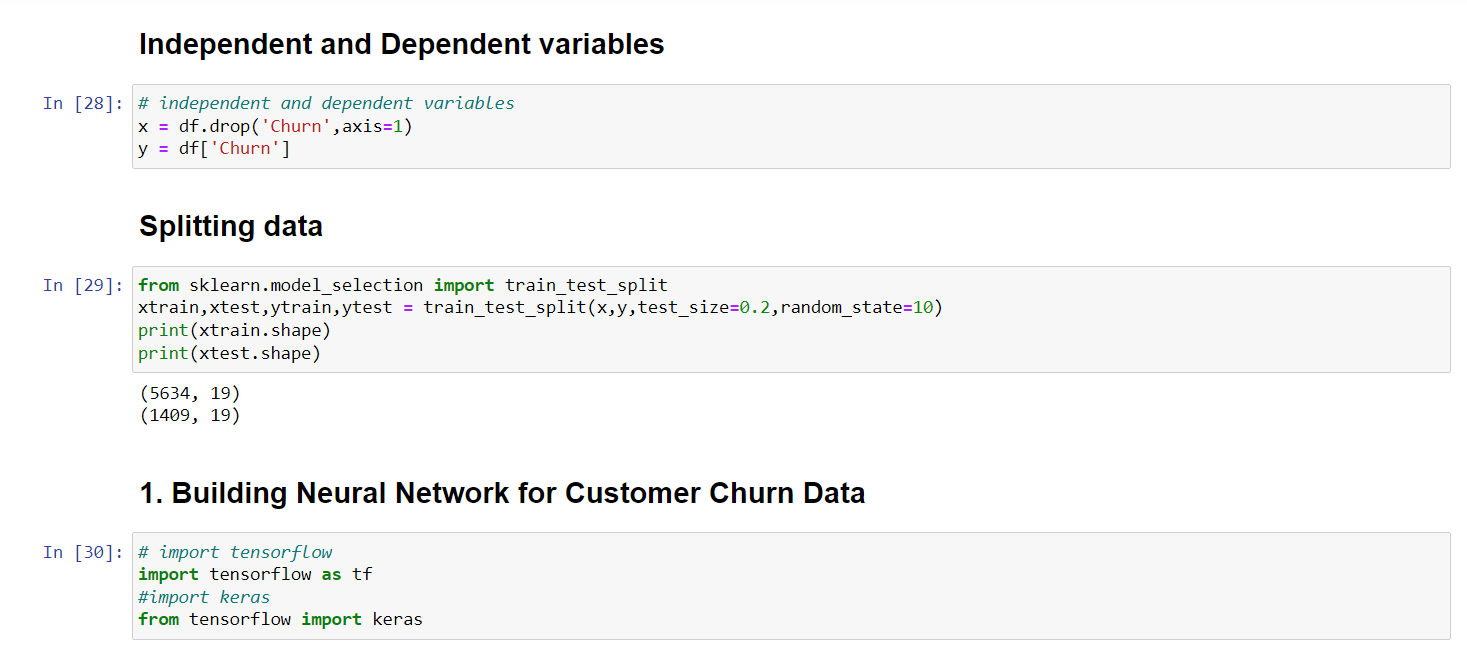
**9. Data Visualization in IBM Cognos Analytics**

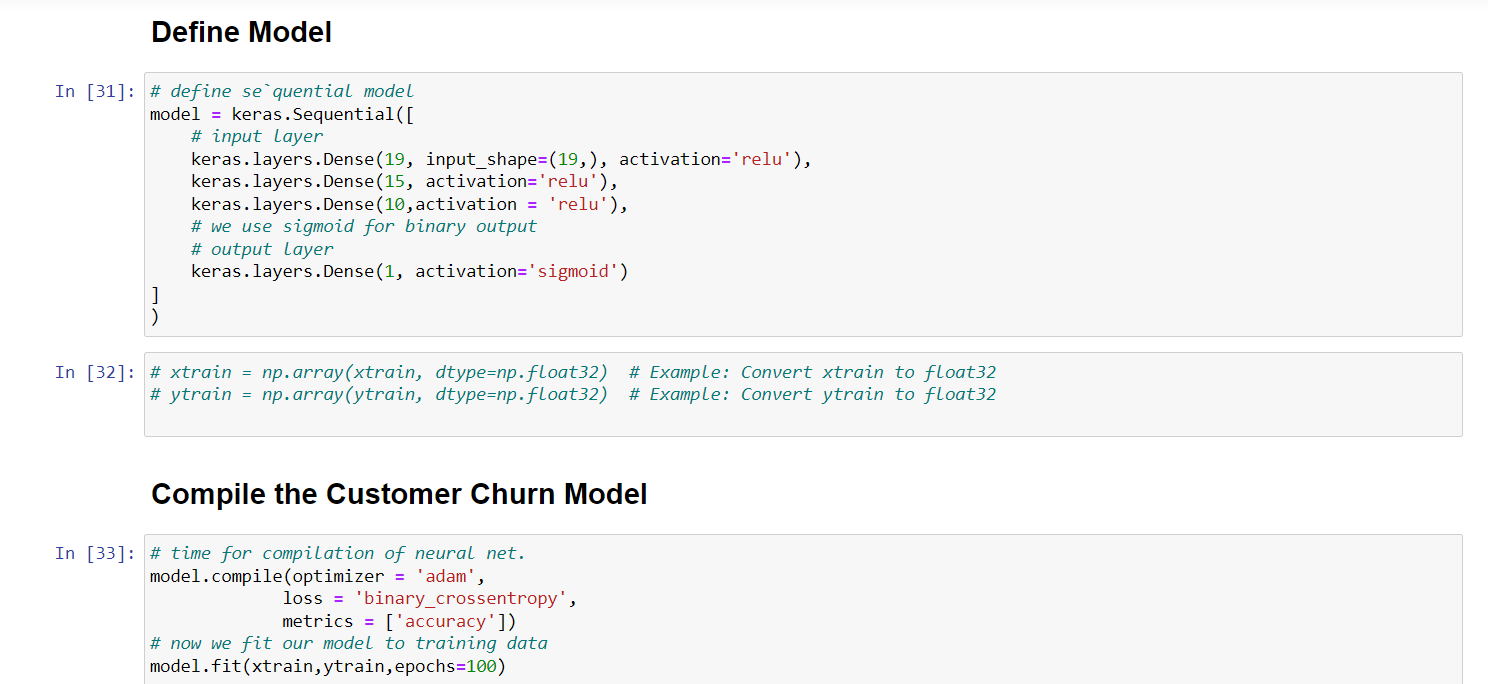


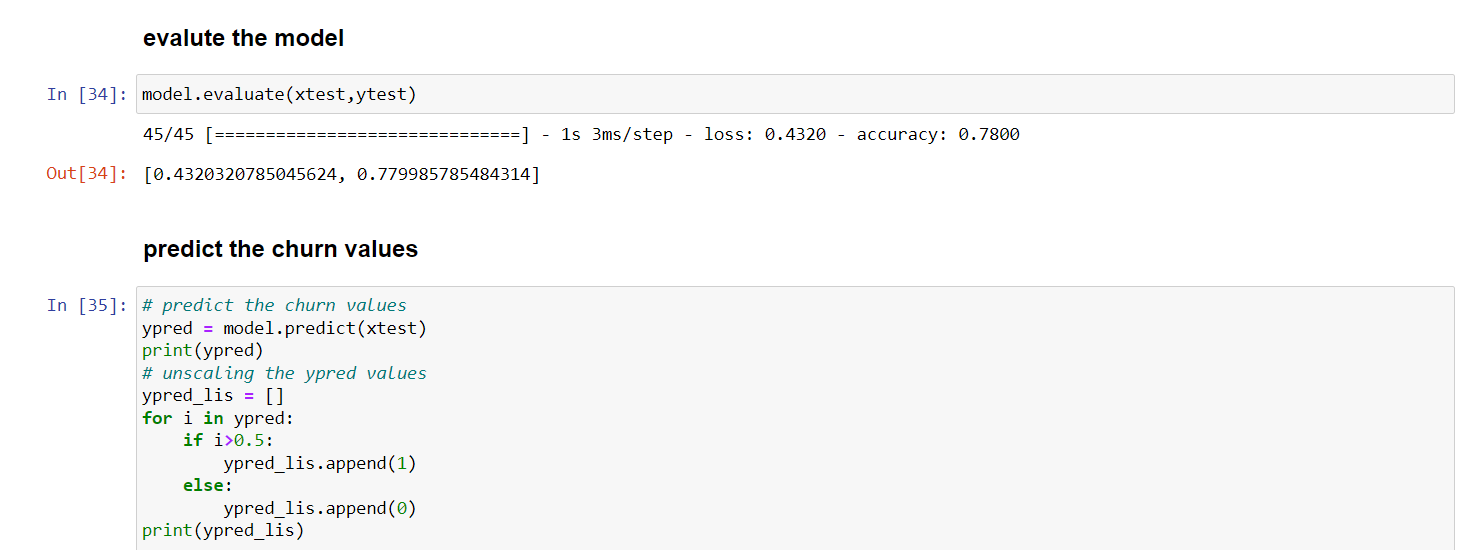


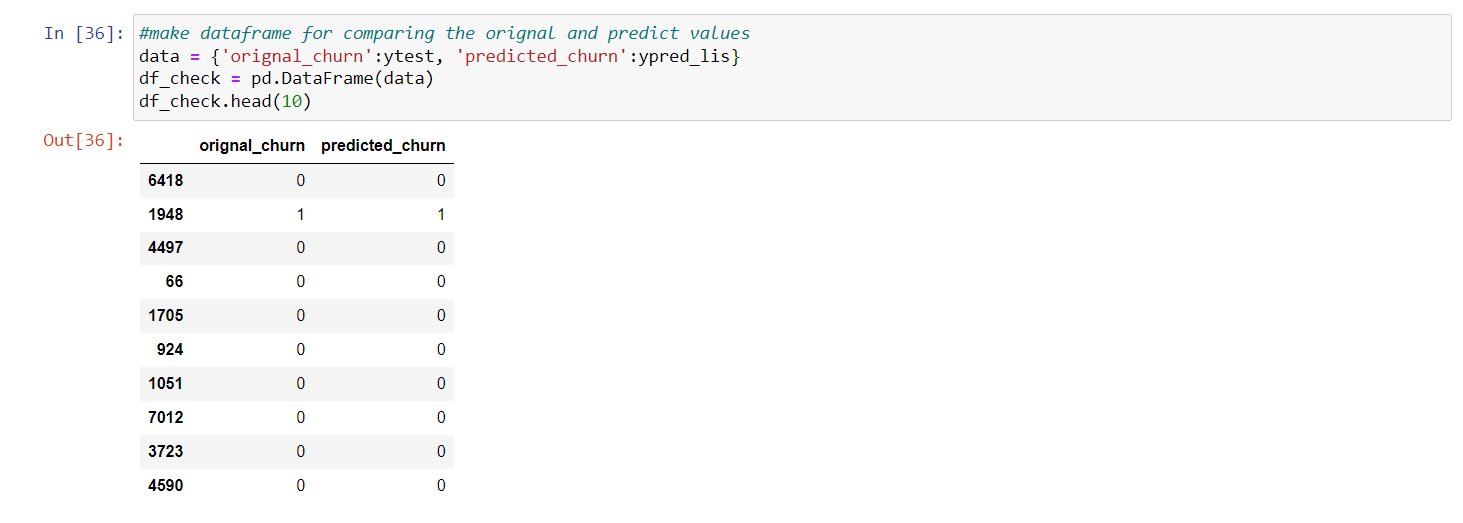
**9.PREDICTIVE MODEL**

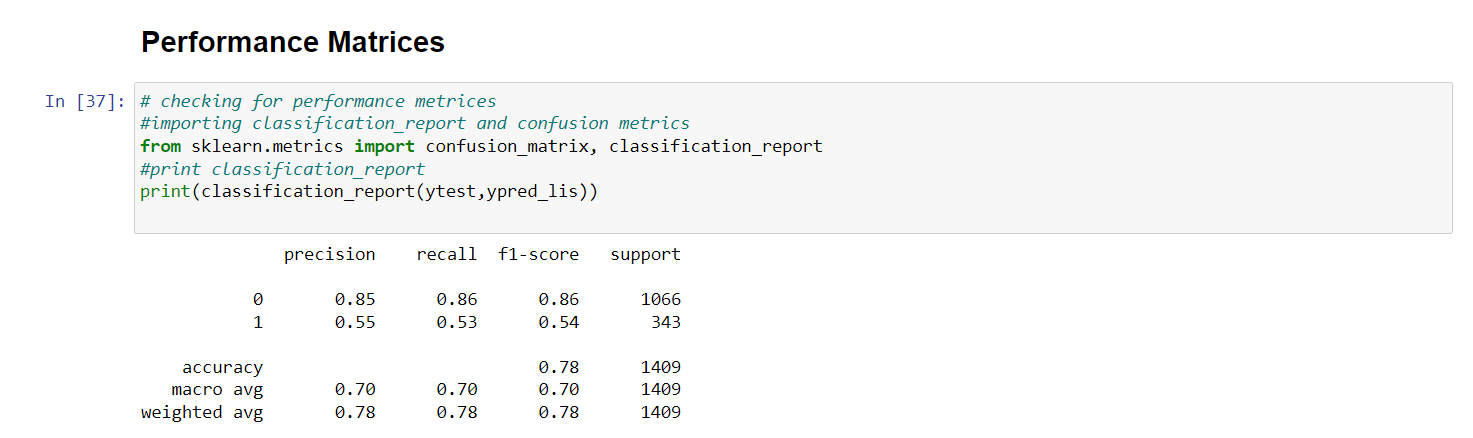
1. Building Neural Network for Customer Churn Data
2. Decision Tree Classifier
3. Random Forest Classifier
4. Logistic Regression
5. Support Vector Machine (SVM)

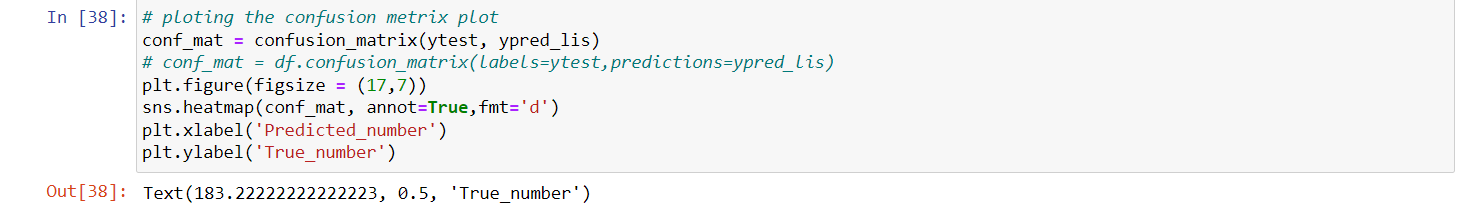


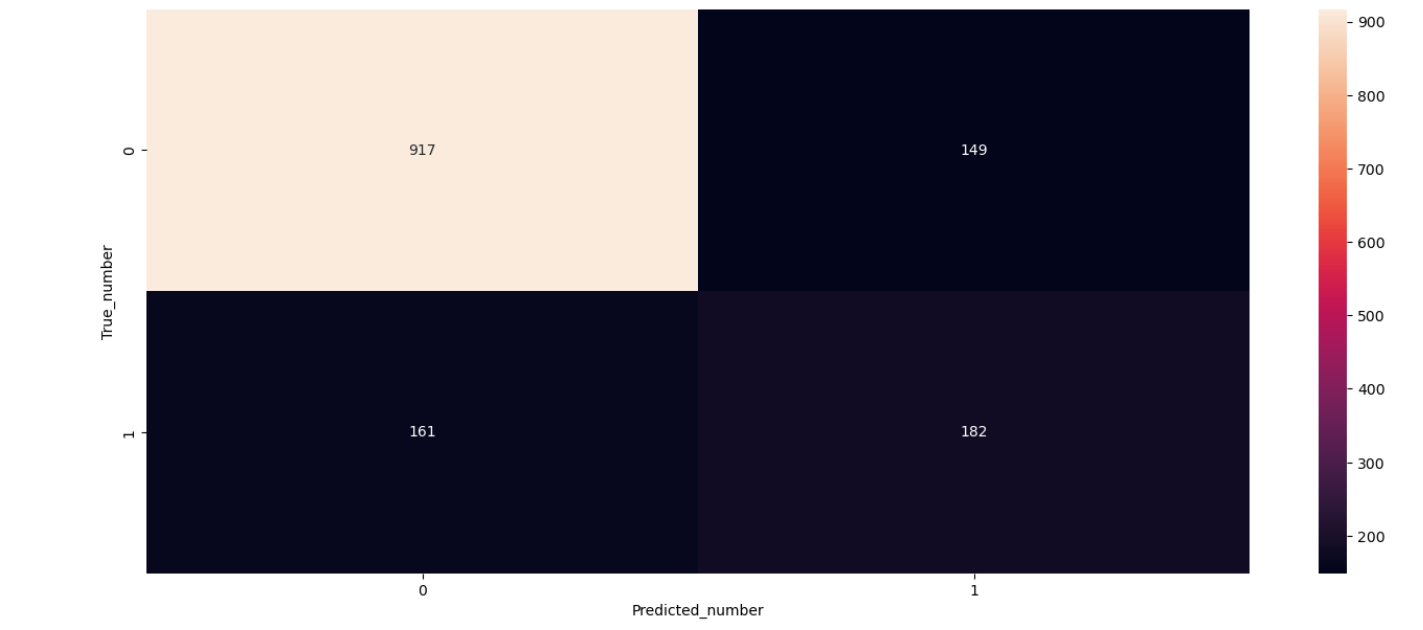






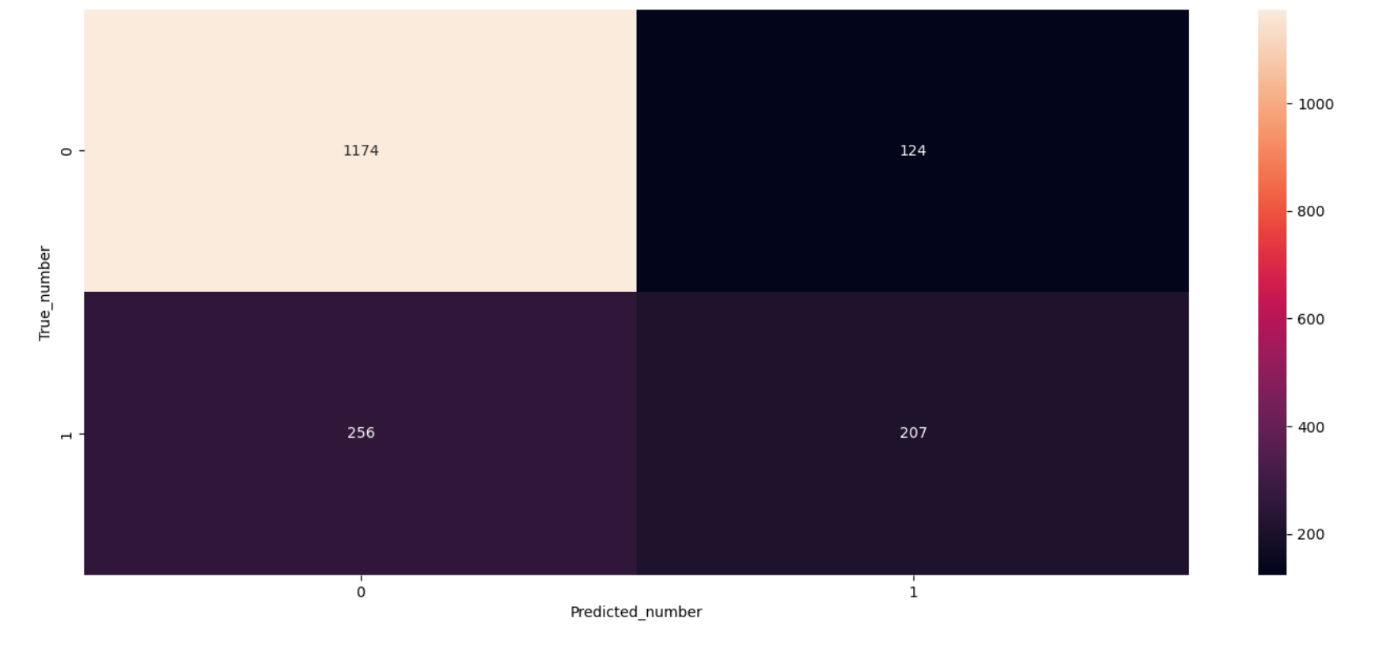


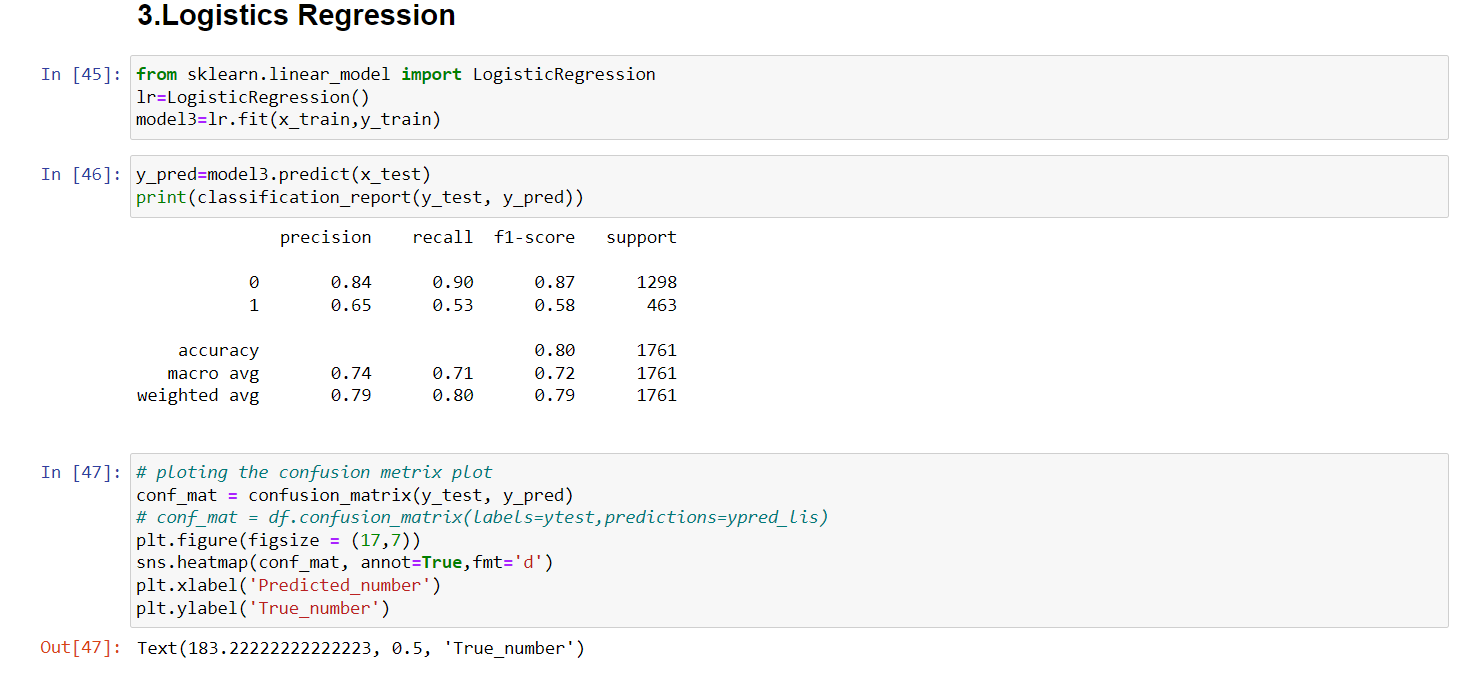


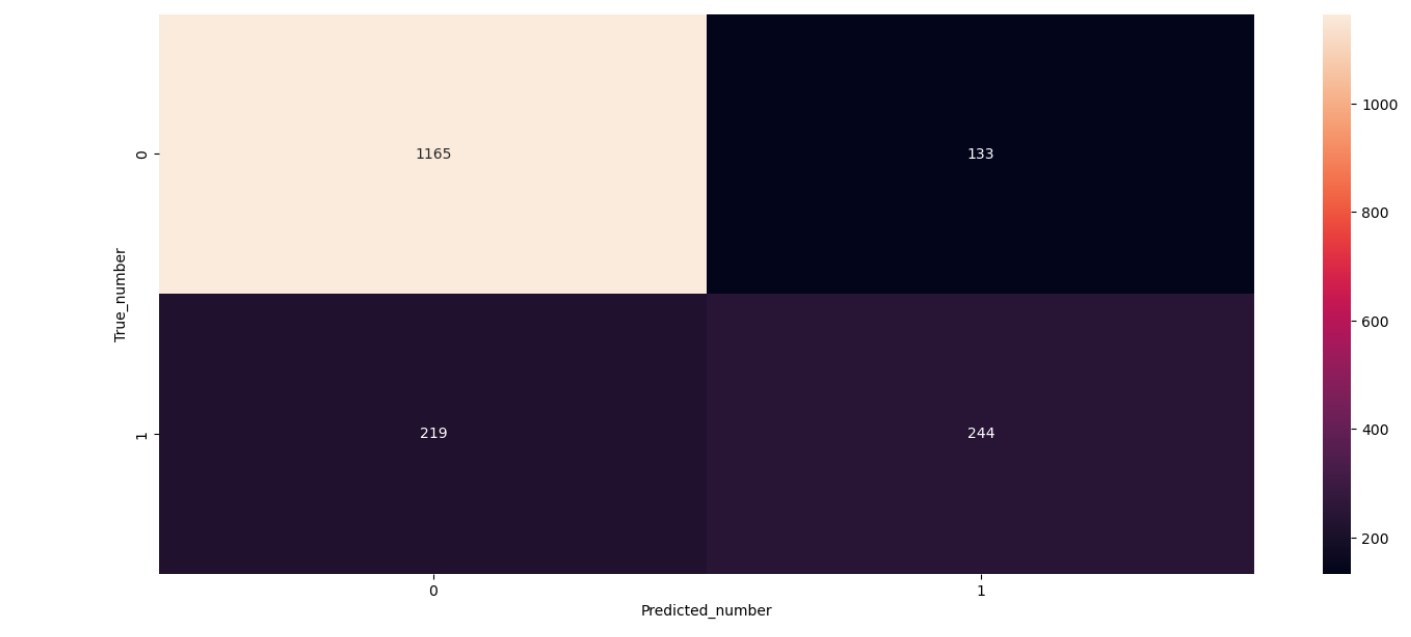


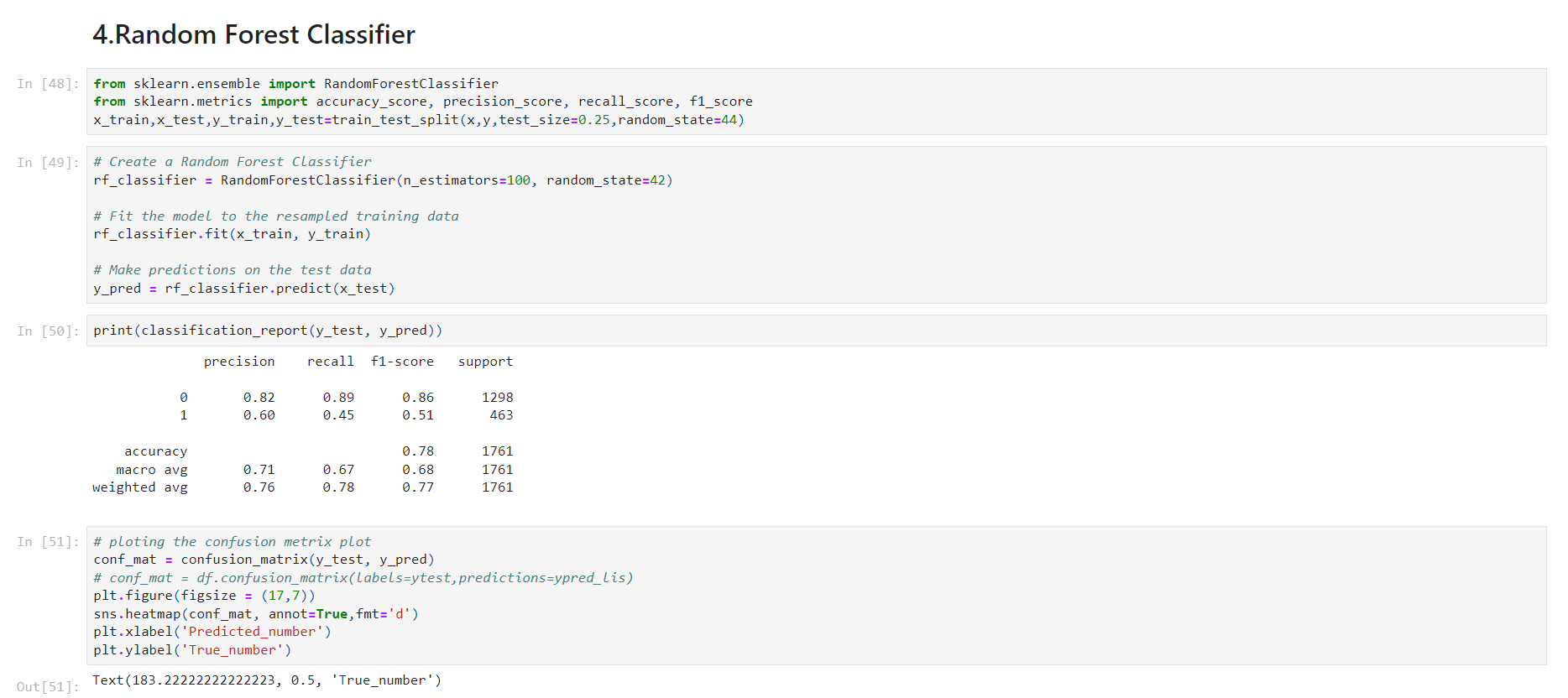


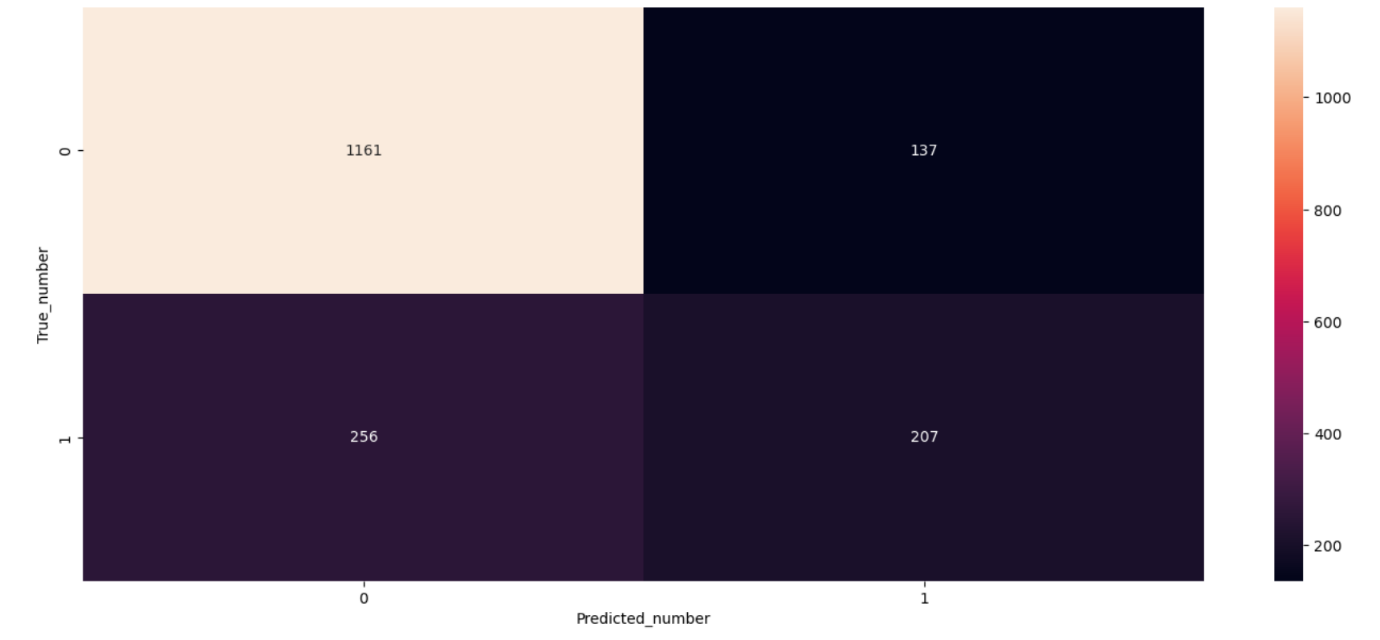


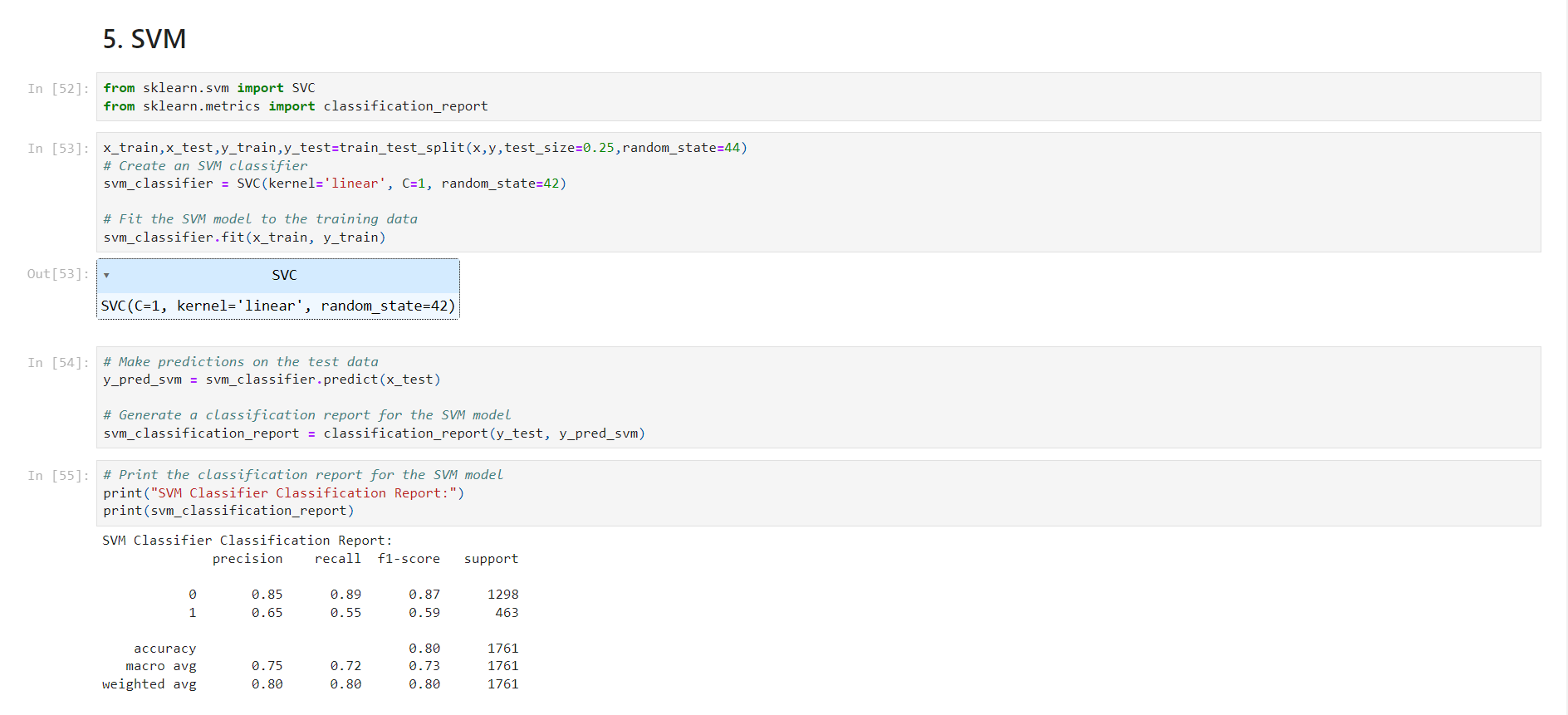


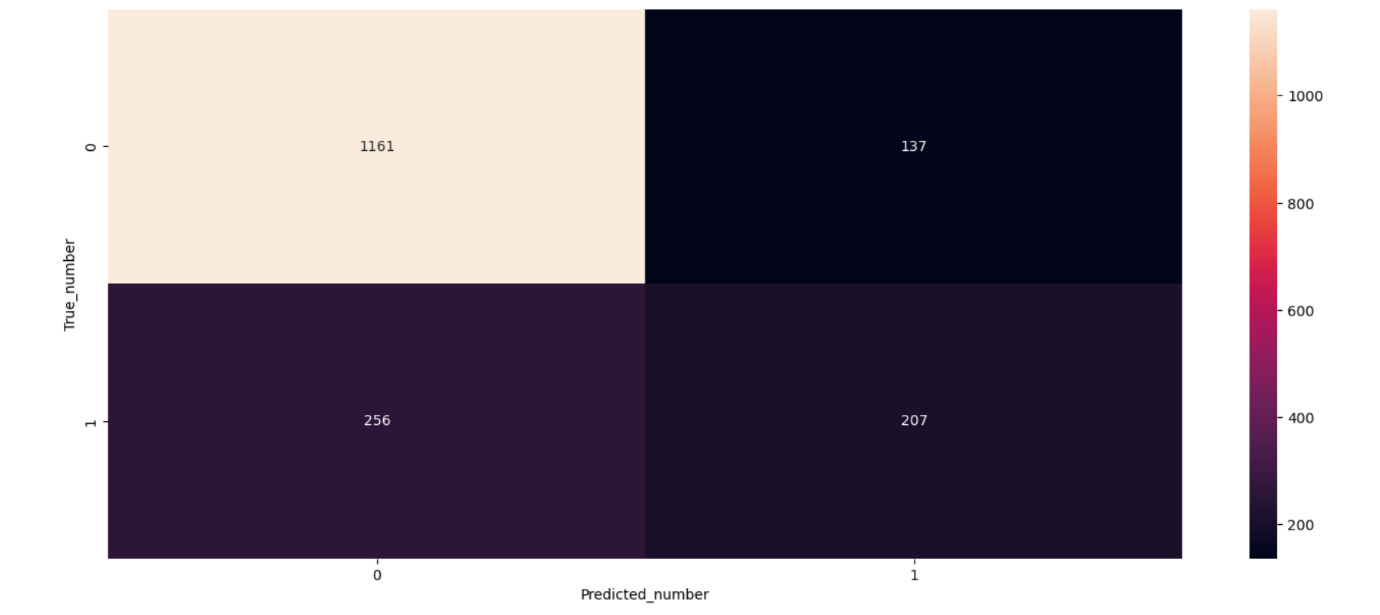




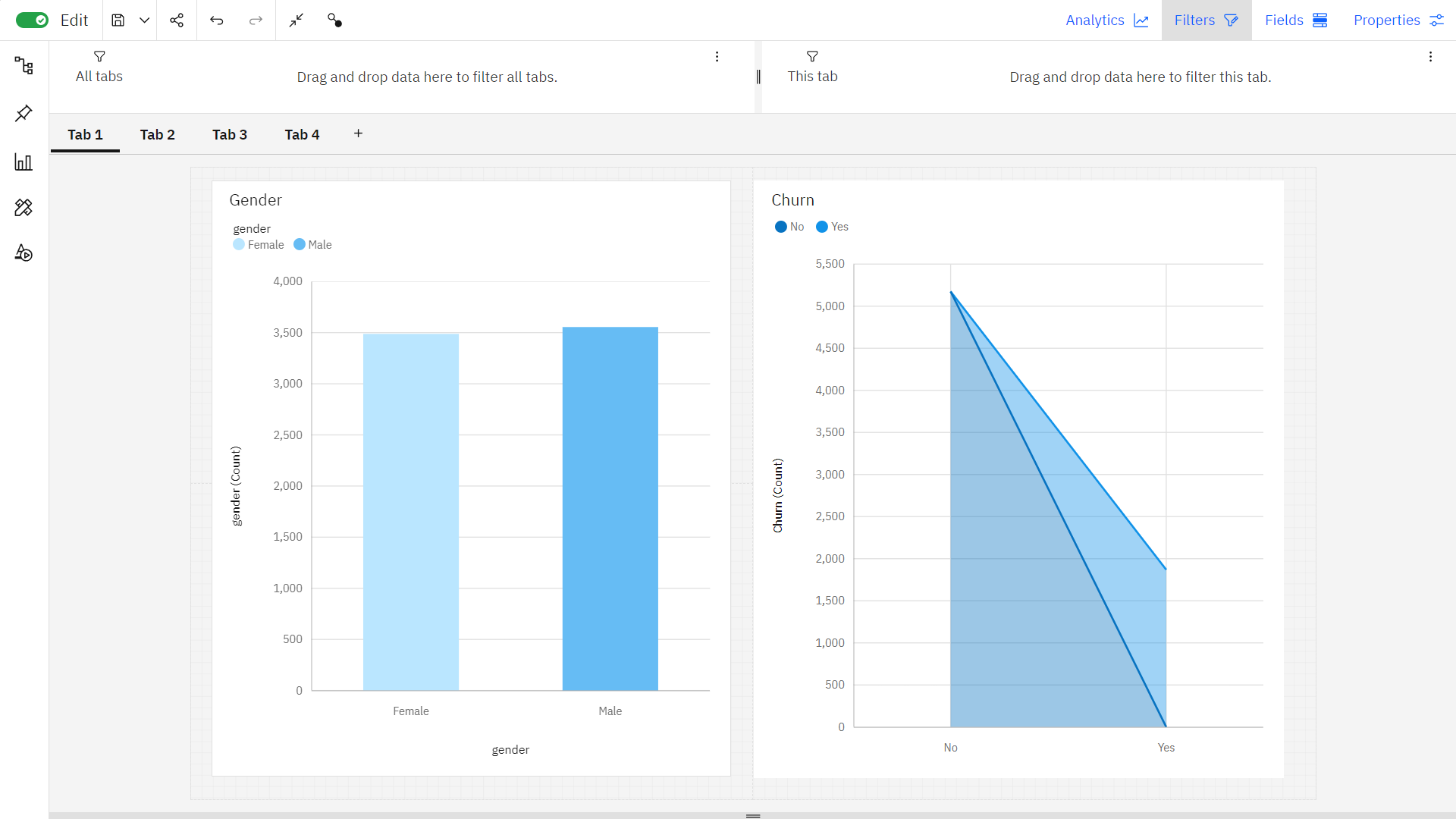




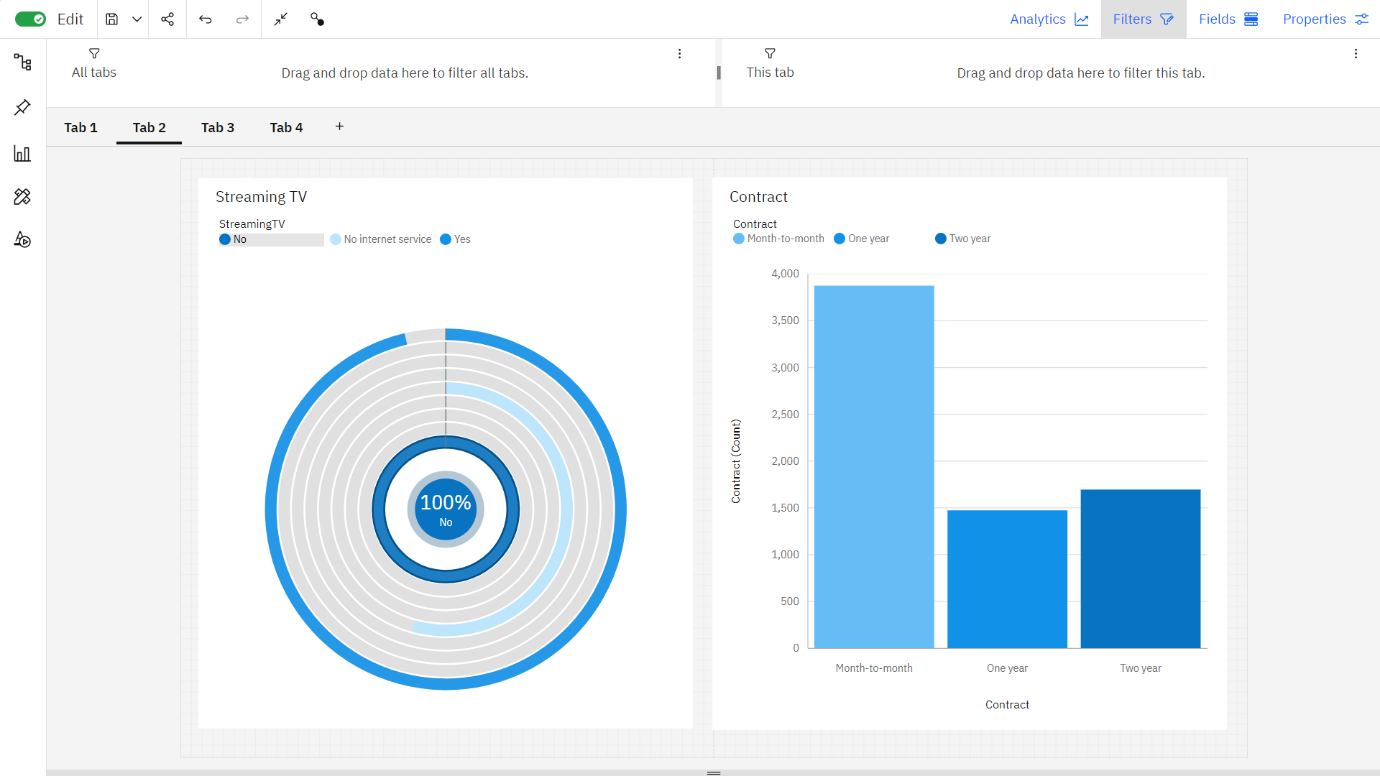




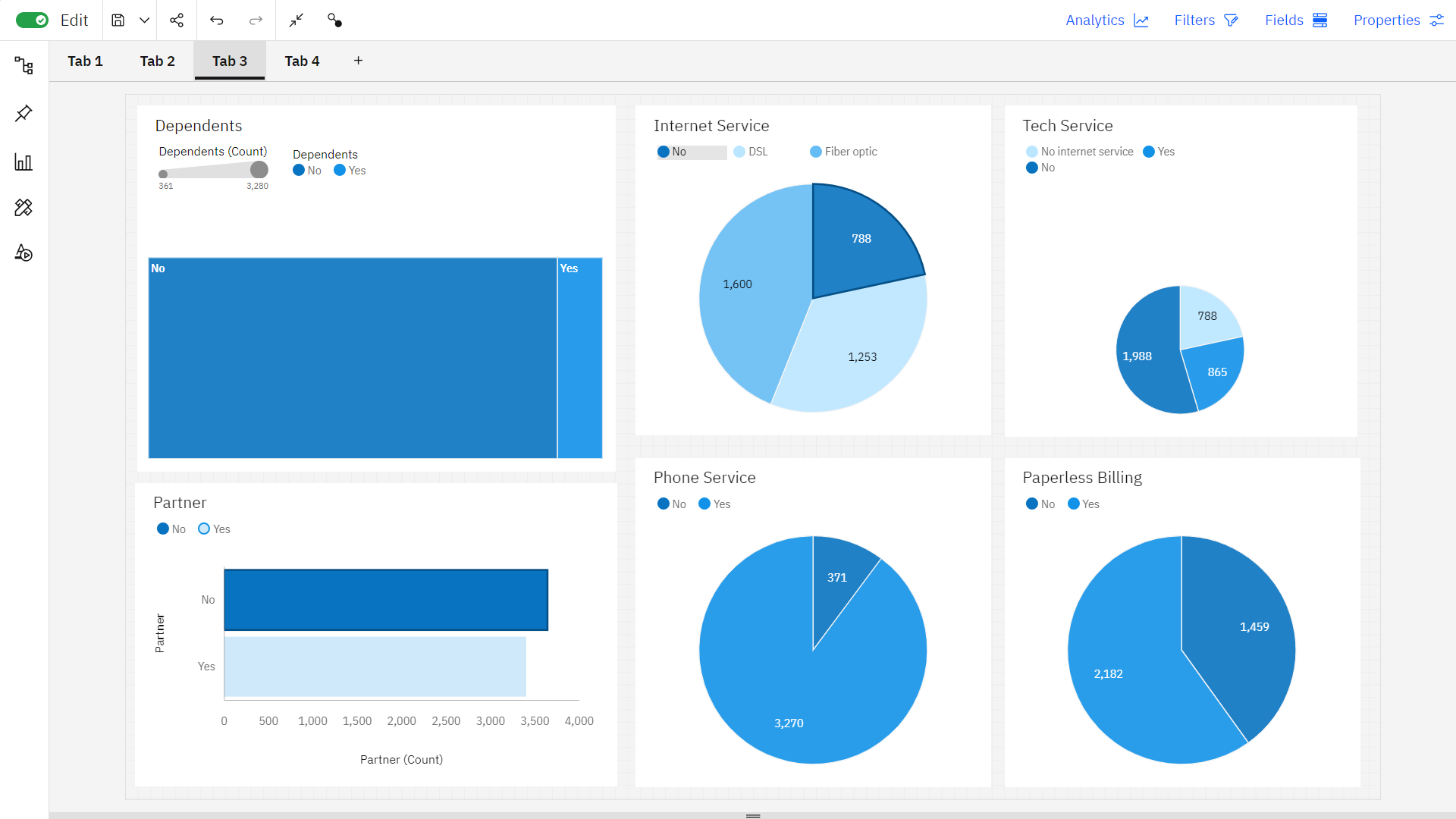
**Dashboard and Report in IBM Cognos Analytics:**



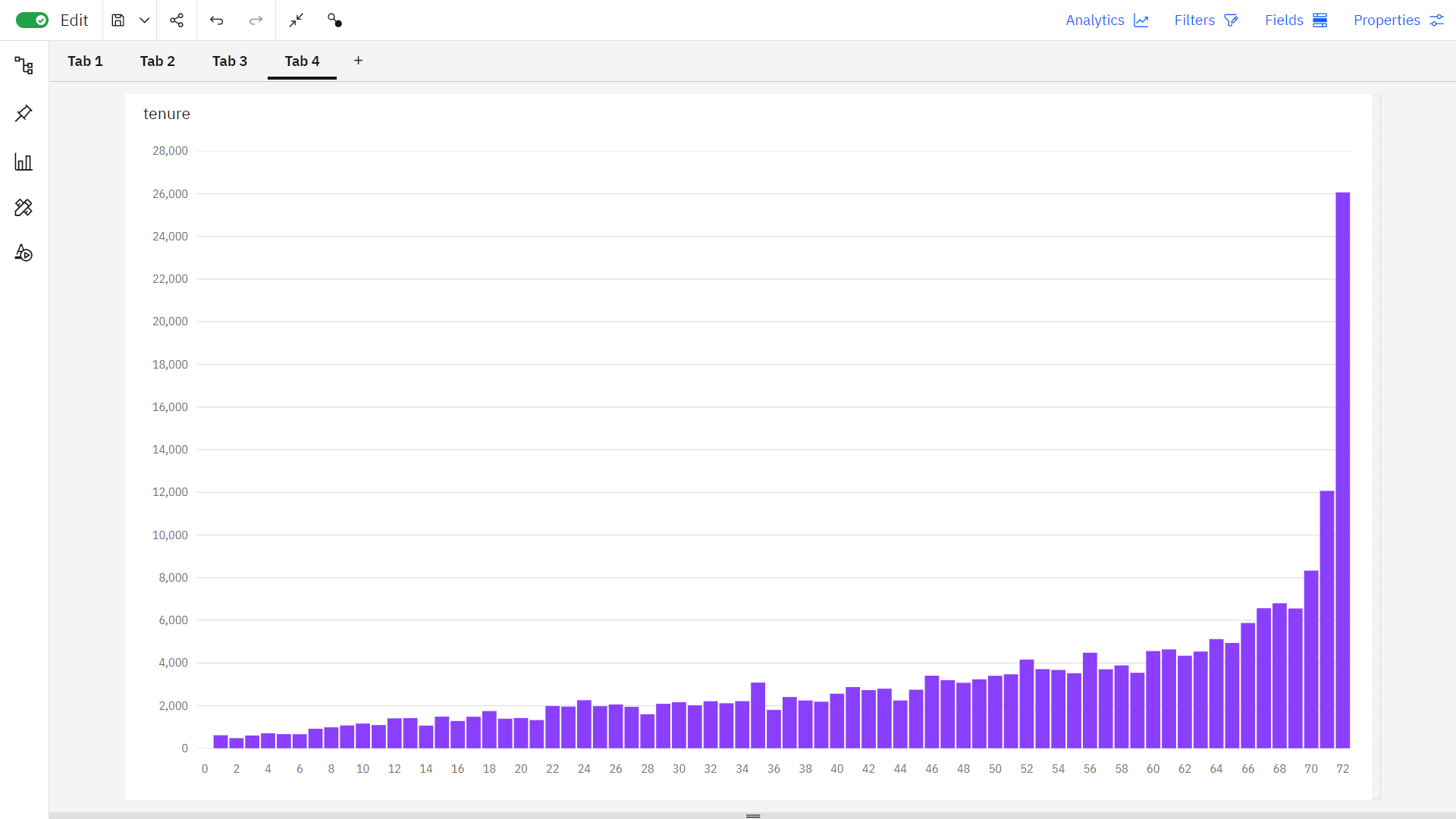
In tab 1, first graph is bar graph which represents gender. It shows Male is the most frequently occurring category of gender with a count of 3555 items with gender values (50.5 % of the total). Second graph is area graph of churn, which shows **No** is the most frequently occurring category of **Churn** with a count of **2549** items with **Churn** values (**73.1** % of the total).



In tab 2, first graph is radial graph of Streaming TV. The total number of results for Streaming TV is over seven thousand. Second graph is stacked column chart of contract duration in which month-to-month is the most frequently occurring category (55 % of the total).



In tab 3, No is the most frequently occurring category of Dependents. In pie chart of Internet service, the total number of results for Internet Service is nearly two thousand. In tech service the count is unusually high when Tech Support is No. The bar chart of total number of results for Partner, across all partners, is over seven thousand. Yes, exceeds No in Phone Service by 1. Yes, exceeds No in Paperless Billing by 1.



In tab 4, it is a Stack column chart which represents the most significant values of tenure are 72, 71, 70, 69, and 68, whose respective tenure values add up to 350, or 13.3 % of the total.

**How the Insights and Predictive Model help the Business reduce Customer Churn?**

**Insights** can help businesses understand why customers are churning. Once businesses understand the reasons for churn, they can develop targeted retention strategies to address those reasons. For example, if customers are churning because they are not satisfied with the product, businesses can improve the product or develop new features that meet the needs of their customers.

**Prediction models** can help businesses identify customers who are at risk of churning. This allows businesses to proactively reach out to these customers and offer them incentives to stay, such as discounts or personalized support.

Specific ways that insights and prediction models can help businesses reduce customer churn:

* **Identify the key factors that are driving churn:** Using insights from customer data, businesses can identify the key factors that are causing customers to churn. This information can then be used to develop targeted retention strategies. For example, if businesses find that customers are churning because they are unhappy with the price of their service, they can develop new pricing plans that are more competitive.
* **Improved customer satisfaction:** By proactively addressing customer concerns and offering them personalized support, businesses can improve customer satisfaction. This can lead to increased customer loyalty and reduced churn.
* **Segment customers based on their churn risk:** Using prediction models, businesses can segment customers based on their churn risk. This allows businesses to focus their retention efforts on the customers who are most likely to churn. For example, businesses can target customers with high churn risk with personalized outreach campaigns or special offers.
* **Reduced costs:** Customer churn can be a costly problem for businesses. By reducing churn, businesses can save money on marketing and acquisition costs.
* **Proactively reach out to customers at risk of churning:** Once businesses have identified customers who are at risk of churning, they can proactively reach out to them to offer support or incentives to stay. For example, businesses can offer discounts to customers who are considering churning.
* **Increased revenue:** By retaining existing customers, businesses can increase their revenue. Existing customers are more likely to purchase additional products or services, and they are also more likely to refer new customers to the business.

**10.CONCLUSION**

Within the dynamic and competitive telecom industry, our customer churn prediction model, leveraging Support Vector Machine (SVM) and Logistic Regression, has proven its worth by providing a higher accuracy rate. This demonstrates its vital role in helping telecom companies identify potential churners and institute proactive strategies for customer retention. The ability of both SVM and Logistic Regression to perform equally well underscores their adaptability in a telecom-specific context. High-quality data remains a linchpin for sustained model accuracy, and continuous model refinement is crucial for staying ahead of evolving customer preferences and industry trends.