PHASE-5

DATA ANALYST WITH COGNOS

PROJECT- CUSTOMER CHURN PREDICTION IN TELECOMMUNICATION

411421104033

MUHAGANI .M

Introduction:

Telecommunication Churn Prediction is a machine learning project which focuses on predict customer churn in the telecommunication industry. Churn prediction is important for telecommunication companies because it helps them to retain their customers and reduce customer acquisition costs. The figure below depicts the description of the data set importation and screenshot of same.

DATA COLLECTION :

In the first step of context of telecommunications or any other industry, data collection is typically Typically performed using other specialized tools and systems. Here's an overview of how data collection in telecommunications works, with Cognos playing a role in data analysis and reporting not used for directcollection of data in telecommunications or any other industry.Data collection in telecommunications involves a complex process of gathering, storing, and processing data from various sources, and specialized tools and systems are used for each stage of this process. Cognos comes into play after the data has been collected and prepared for analysis, providing a means to visualize and report on the insights gained from the data.

Data Source:

Telecommunication companies collect a data from a wide range of sources includes a networks equipment,call detail records,customer relationship management system,billing system,network monitoring tools .

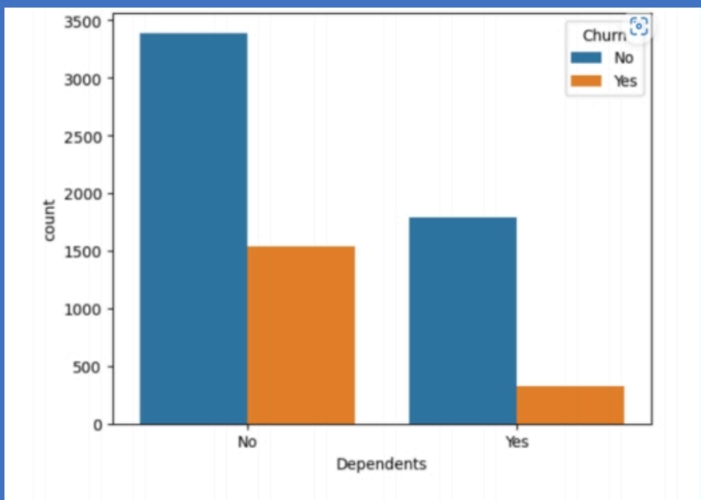
Dataset link : (https://www.kaggle.com/datasets/blastchar/telco-customer-churn)

Customer ID	gender	Senior Citizen	Partner	Dependents	tenure	Phone Service	Multiple Lines
7590-VHVEG	Female	0	Yes	No	1	No	No phone service
5575-GNVDE	Male	0	No	No	34	Yes	No
3668-QPYBK	Male	0	No	No	2	Yes	No
7795-CFOCW	Male	0	No	No	45	No	No phone service
9237-HQITU	Female	0	No	No	2	Yes	No
9305-CDSKC	Female	0	No	No	8	Yes	Yes
1452-KIOVK	Male	0	No	Yes	22	Yes	Yes
6713-OKOMC	Female	0	No	No	10	No	No phone service
7892-POOKP	Female	0	Yes	No	28	Yes	Yes
6388-TABGU	Male	0	No	Yes	62	Yes	No
9763-GRSKD	Male	0	Yes	Yes	13	Yes	No
7469-LKBCI	Male	0	No	No	16	Yes	No
8091-TTVAX	Male	0	Yes	No	58	Yes	Yes
0280-XJGEX	Male	0	No	No	49	Yes	Yes
5129-JLPIS	Male	0	No	No	25	Yes	No
3655-SNQYZ	Female	0	Yes	Yes	69	Yes	Yes
8191-XWSZG	Female	0	No	No	52	Yes	No
9959-WOFKT	Male	0	No	Yes	71	Yes	Yes
4190-MFLUW	Female	0	Yes	Yes	10	Yes	No
4183-MYFRB	Female	0	No	No	21	Yes	No
8779-QRDMV	Male	1	No	No	1	No	No phone service
1680-VDCWW	Male	0	Yes	No	12	Yes	No
1066-JKSGK	Male	0	No	No	1	Yes	No
3638-WEABW	Female	0	Yes	No	58	Yes	Yes
6322-HRPFA	Male	0	Yes	Yes	49	Yes	No
6865-JZNKO	Female	0	No	No	30	Yes	No

Churn rate by dependent

```
def target_vs_category_visual(dataframe,target, categorical_col):
    plt.figure(figsize=(15,8))
    sns.histplot(x=target,hue=categorical_col, data=dataframe,element="step",multiple="dodge")
    plt.title("State of Categorical Variables according to Churn ")
    plt.show()

for col in cat_cols:
    target_vs_category_visual(df,"Churn",col)
```



Those without dependents tend to churn more than those with dependents.

Model Evaluation and Prediction :

Model evaluation is a critical step in the telecommunication prediction process. It helps assess the performance and accuracy of the prediction models we've developed. It can splitting the data, Selecting evaluation metrics and Some models are logistic regression Decision Trees Support vector machine(SVM),Neural Networks),Model Training and Validation, Visualization and plots, Comparative Analysis and Communication results .Effective model evaluation ensures that your telecommunication prediction models are accurate, reliable, and aligned with the goals of your organization.

Model Prediction Predictive modeling in telecommunication can be used for various purposes, such as predicting customer churn, forecasting network traffic, optimizing resource allocation, and improving customer satisfaction. It allows telecom companies to make data-driven decisions and respond proactively to industry changes and customer needs.

PREDICTIVE MODELLING :

Predictive modeling in the field of telecommunications involves using historical data and statistical algorithms to make predictions about future events or outcomes. Data analysts and data scientists in the telecommunications industry can leverage predictive modeling techniques to address various challenges and opportunities

PYTHON PROGRAM:

```
#Import necessary libraries

# Function to predict churn based on inputs

def predict_churn(customer_id, gender, senior_citizen, partner, dependents, phone_service, tenure, multiple_lines):

  # You would typically use a machine learning model for such predictions

  # This is a simplified example and does not use a real model


  # Convert binary values (yes/no) to 1/0

  partner = 1 if partner.lower() == "yes" else 0
```

```
    dependents = 1 if dependents.lower() == "yes" else 0
    phone_service = 1 if phone_service.lower() == "yes" else 0
    multiple_lines = 1 if multiple_lines.lower() == "yes" else 0

    # Example of a simple rule-based prediction
    if (senior_citizen == 1 and tenure < 12) or (partner == 0 and dependents == 0 and tenure < 6):
        churn_prediction = "Yes"
    else:
        churn_prediction = "No"

    # Print the prediction
    print("Customer ID:", customer_id)
    print("Churn Prediction:", churn_prediction)

# Input values (you can replace these with user input or actual data)
customer_id = "9237-HQITU"
gender = "Female"
senior_citizen = 0  # 1 for yes, 0 for no
partner = "Yes"  # Yes or No
dependents = "No"  # Yes or No
phone_service = "Yes"  # Yes or No
tenure = 2  # Number of months
multiple_lines = "No"  # Yes or No

# Call the predict_churn function with input values
predict_churn(customer_id, gender, senior_citizen, partner, dependents, phone_service, tenure, multiple_lines)
```

OUTPUT:

```
Customer ID: 9237-HQITU
Churn Prediction: No
```

Data Analysis

❖ After cleaning the data set, the next step was to perform exploratory data analysis (EDA). EDA is important because it helps to understand the distribution of the data and identify any trends or patterns in the data.

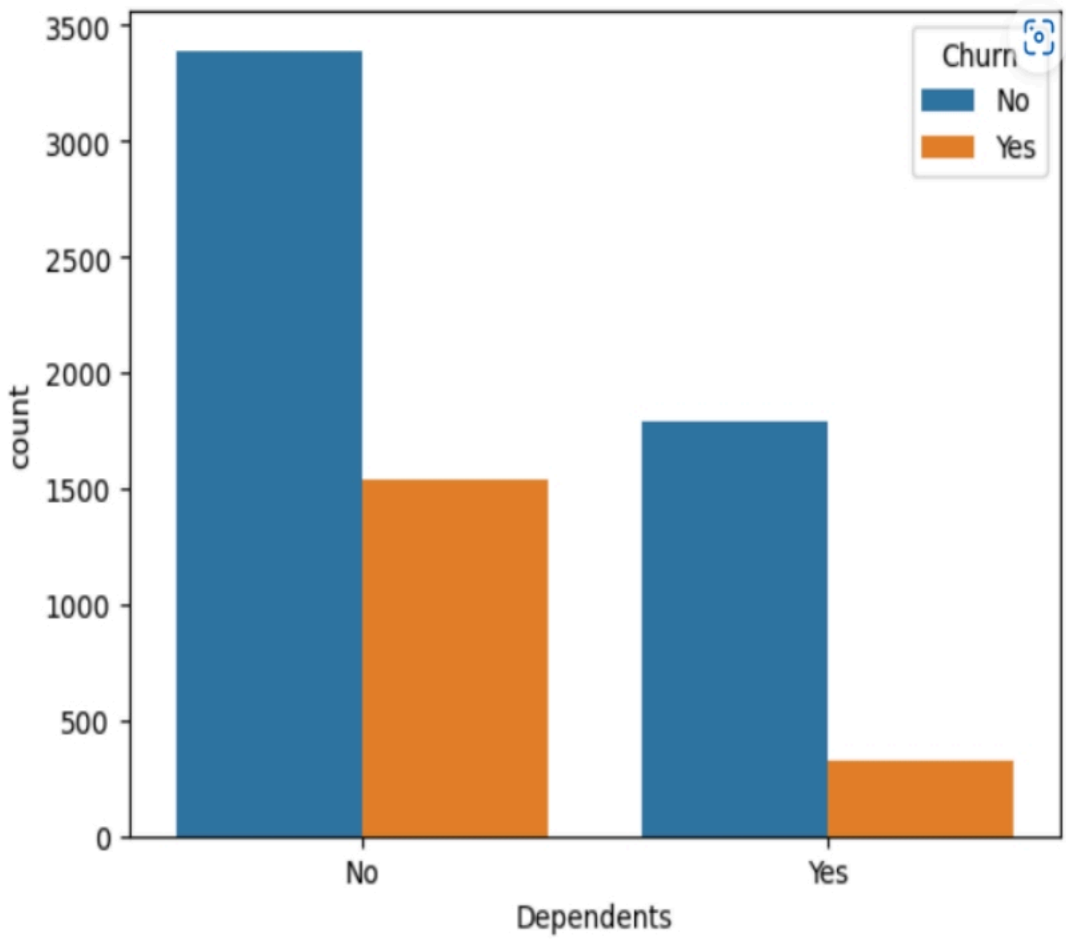
- ❖ The first step in EDA was to analyze the distribution of gender of the target variable ‘Churn’. it was revealed the distribution is almost equally with a difference of 67 count of male than female.
- ❖ The ratio between the Senior Citizens to Non-Senior Citizens was about 1:5. With Non-Senior Citizens constituting about 84% of the whole distribution. Thus, most of the customers in the dataset are younger people.
- ❖ About 50% of the customers have a partner, while only 30% of the total customers have dependents.
- ❖ Interestingly, among the customers who have a partner, only about half of them also have a dependent, while other half do not have any dependents.
- ❖ Additionally, as expected, among the customers who do not have any partner, a majority (90%) of them do not have any dependents.

Overall Churn Rate

In our data, 74% of the customers do not churn. Clearly the data is skewed as we would expect a large majority of the customers to not churn. This is important to keep in mind for our modelling as skewed-sens could lead to a lot of false negatives. We will see in the modelling section on how to avoid newness in the data.

1 Explore the churn rate by dependent, seniority, payment method, monthly charges and total charges to see how it varies by these variables.

A. Churn rate by dependent

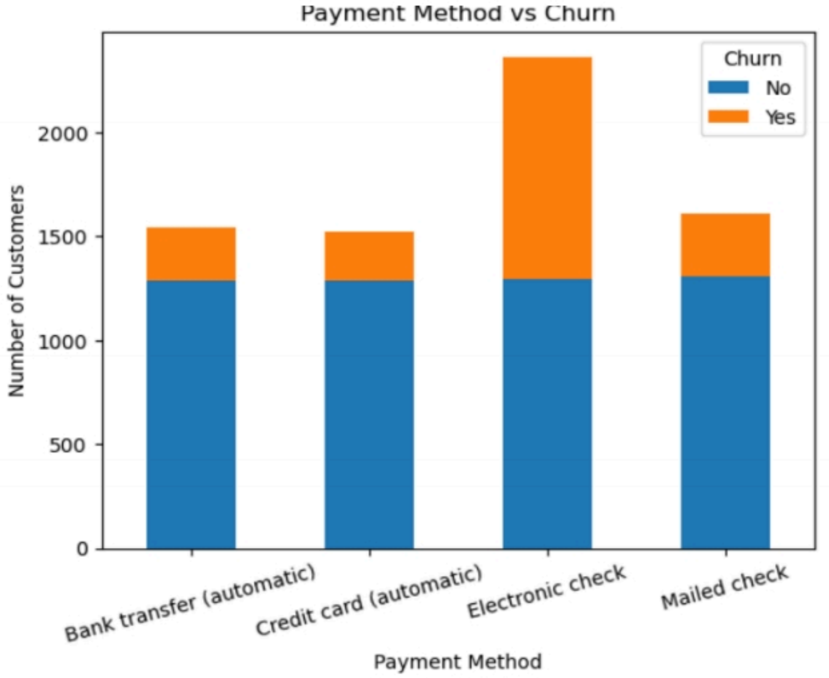


Those without dependents tend to churn more than those with dependents.

Churn Rate By Dependent

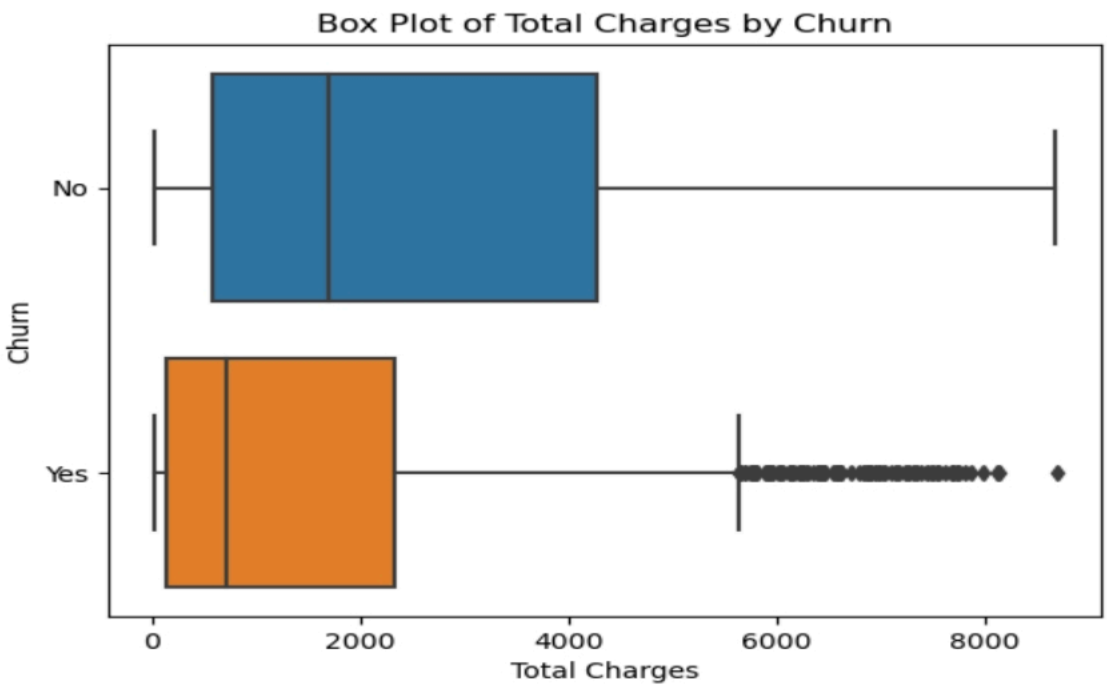
As we can see from the above plot, the customers who do not have dependent tend to churn more than those with dependents.

B. Churn by payment method



2

C. Churn by total charges



Machine learning Modeling:

After performing EDA, the next step was to build a machine learning model to predict customer churn. The following models were used for this project; Logistic Regression, Random Forest, Gradient Boosting, Neural Network, Support Vector Machine (Linear Kernel), Support Vector Machine (RBF Kernel), K-Nearest Neighbors, Decision Tree.

	Model	Precision_Score	Recall_Score	F1_Score	Accuracy
0	Logistic Regression	0.656716	0.592593	0.623009	0.811003
1	Gradient Boosting	0.688797	0.558923	0.617100	0.817214
2	Support Vector Machine (Linear Kernel)	0.649606	0.555556	0.598911	0.803904
3	Random Forest	0.644351	0.518519	0.574627	0.797693
4	Support Vector Machine (RBF Kernel)	0.681159	0.474747	0.559524	0.803017
5	K-Nearest Neighbors	0.546667	0.552189	0.549414	0.761313
6	Neural Network	0.586207	0.515152	0.548387	0.776398
7	Decision Tree	0.491582	0.491582	0.491582	0.732032

The 8 models were trained on the training data and evaluated on the test data using accuracy, Precision, Recall, and F1-score metrics. The Logistic Regression model had the highest performance with an F1 score of 0.62. The top four performing models were selected and tuned.

Summary of Results:

	model	F1_score	Accuracy
0	Logistic Regression	0.723982	0.801637
1	Support Vector Machine (Linear Kernel)	0.715923	0.795870
2	Gradient Boosting	0.715420	0.794092
3	Neural Network	0.712535	0.794981

- Based on the Accuracy and F1-Score metrics, the best model for the classification problem appears to be Logistic Regression model. Although all the other models had similar accuracy scores, Logistic Regression model had the highest F1-Score and Accuracy.

- While Logistic Regression model had a good accuracy score of 0.80, its F1-Score, is slightly higher than the other three models. Therefore, based on the metrics evaluated Logistic Regression model appears to be the best model for this classification problem.

- ❖ However, it is important to note that other factors such as model complexity, computerization, efficiency, and interpret-ability such also be considered when choosing a model for deployment.

Conclusion

In conclusion, the integration of data analytics with Cognos in the field of telecommunications holds significant potential for transforming the industry. Data analytics, when harnessed effectively, can provide telecom companies with valuable insights that can lead to improved operational efficiency, enhanced customer experiences, and informed decision-making. Some key takeaways from the utilization of data analytics with Cognos in telecommunications include:

1. **Enhanced Decision-Making:** Telecom companies can make more informed decisions by analyzing vast amounts of data related to network performance, customer behavior, and market trends. This can lead to better resource allocation and strategic planning.
2. **Revenue Growth:** Data analytics can uncover opportunities for upselling, cross-selling, and more effective marketing strategies, thereby increasing revenue for telecom companies.
3. **Cost Reduction:** By identifying areas of inefficiency and optimizing resource allocation, data analytics can help telecoms reduce operational costs.
4. **Compliance and Security:** Telecoms can use data analytics to monitor and maintain compliance with regulatory requirements and enhance cybersecurity measures to protect sensitive customer data.

However, it's important to note that the successful implementation of data analytics with Cognos in the telecommunications sector also comes with challenges, such as data privacy concerns, data quality issues, and the need for skilled data analysts and data scientists.the combination of data analytics with Cognos has the potential to revolutionize the telecommunications industry, providing companies with a competitive edge and enabling them to deliver better services, reduce costs, and drive innovation. It is crucial for telecom organizations to invest in the necessary technology and talent to harness the full potential of data analytics and Cognos in their operations.