Customer Churn

Mustapha Hadrous

EC Utbildning

Project Data Science

19830102

Abstract

In today's competitive business environment, retaining customers is crucial for long-term success. Churn analysis is a key technique used to understand and reduce this customer attrition. It involves examining customer data to identify patterns and reasons behind customer departures. By using advanced data analytics and machine learning, businesses can predict which customers are at risk of leaving and understand the factors driving their decisions. This knowledge allows companies to take proactive steps to improve customer satisfaction and loyalty.

This project is helping to understand weakness and predict future possibilities of Churn.

Förkortningar och Begrepp

- Churn: is a measurement of the percentage of accounts that cancel or choose not to renew their subscriptions

- SQL processing: is the parsing, optimization, row source generation, and execution of a SQL statement

- Preparing model in machine learning: building a machine learning model involves collecting and preparing data, selecting the right algorithm, tuning it, evaluating its performance, and deploying it for real-time decision-making

- Power BI: is a technology-driven business intelligence tool provided by Microsoft for analysing and visualizing raw data to present actionable information

Skapas automatiskt i Word genom att gå till Referenser > Innehållsförteckning.

Innehållsförteckning

1 Inledning

It involves examining customer data to identify patterns and reasons behind customer departures. By using advanced data analytics and machine learning, businesses can predict which customers are at risk of leaving and understand the factors driving their decisions. This project allows to take steps to improve customer satisfaction and loyalty**.**

Although this project focuses on churn analysis for a telecom firm, the techniques and insights are applicable across various industries. From retail and finance to healthcare and beyond, any business that values customer retention can benefit from churn analysis. We will explore the methods, tools, and best practices for reducing churn and improving customer loyalty, transforming data into actionable insights for sustained success.

2 Teori

Create an entire SQL process in a database & Power BI dashboard to utilize the Customer Data and achieve below goals:

- Visualize & Analyse Customer Data at below levels
- Demographic.
- Geographic.
- Payment & Account Info.
- Services.
- Study Churner Profile & Identify Areas for Implementing Marketing Campaigns.
- Identify a Method to Predict Future Churners.
- Total Customers.
- Total Churn & Churn Rate.
- New Joiners.

2.1 Predict Customer Churn (Using Python)

For predicting customer churn and after testing different Algorithm to get best score , so , we will be using Machine Learning algorithm called **RANDOM FOREST.**

What is Random Forest?

A random forest is a machine learning algorithm that consists of multiple decision trees. Each decision tree is trained on a random subset of the data and features. The final prediction is made by averaging the predictions (in regression tasks) or taking the majority vote (in classification tasks) from all the trees in the forest. This ensemble approach improves the accuracy and robustness of the model by reducing the risk of overfitting compared to using a single decision tree.

Create Churn Prediction Model – Random Forest:

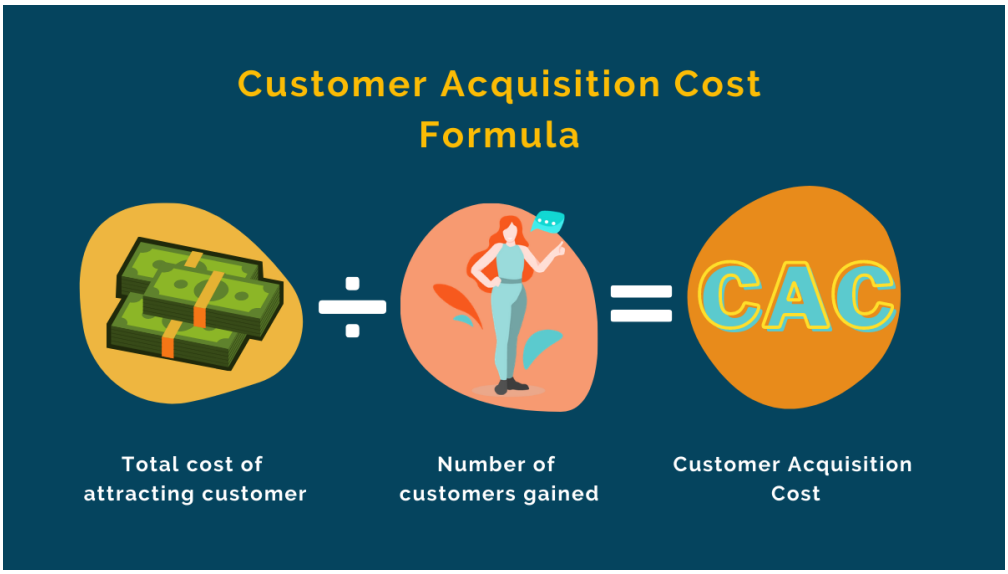
- Importing Libraries & Data Load.
- Data Preprocessing.
- Train Random Forest Model.
- Evaluate Model.
- Save Model for Prediction on New Predicton Data as CSV.
- Logis c Regression: A sta s cal model that esmates the probability of churn based on various independent variables (e.g., customer characteris cs, service usage).

2.2 Metrics for Churn Evaluation:

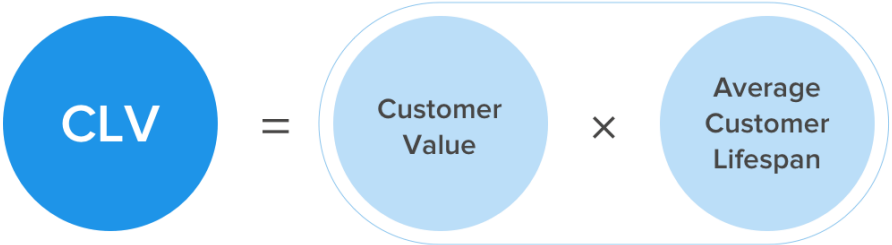
Metrics are defined as a system or standard of measurement. An Acquisition metric is similar, but is far more specific. An acquisition metric is a system or standard of measurement centered around acquiring something. In these Cases, that something is variabel customers :

- **Customer churn rate:** is a metric that shows the total percentage of customers who stop doing business with you over a certain period of time.

- **Customer Acquisition Cost (CAC)**



- **Customer Life Time Value (CLV)**



3 Metod:

Chrun analysis is a method of analyzing customer behavior by grouping customers based on shared characteristics or experiences within a defined timeperiod. In the context of churn analysis is invaluable as it allows businesses to track patterns and trends in customer attrition over time. This method provides a more nuanced understanding of churn, enabling businesses to identify specific periods or factors that influence customer retention.

3.1 Processing Data in SQL Server:

So, the first step in churn analysis is to load the data from our source file. For this purpose, we will be using Microsoft SQL and we do the following:

- Creating Database
- Data Exploration – Check Distinct Value
- Data Exploration – Check Nulls
- Remove null and insert the new data into Production_Churn
- Create two View's for Power BI:
    1. Customer_Status In ('Churned', 'Stayed').
    2. Customer_Status = 'Joined'.

3.2 Power BI Transform:

- Add a new column in production_Churn.
- Create a New Table Reference for Age.
- Create a new measures.
- Create a new table for Productions_Services.

3.3 Power Bi Visualization :

1. Top Card:
    a. Total Customers.
    b. New Joiners.
    c. Total Churn.
    d. Churn Rate%.
2. Demographic:
    a. Gender – Churn Rate.
    b. Age Group – Total Customer & Churn Rate.
3. Account Info:
    a. Payment Method – Churn Rate.
    b. Contract – Churn Rate.
    c. Tenure Group - Total Customer & Churn Rate.

4. Geographic:
    a. Top 5 State – Churn Rate.
5. Churn Distribution:
    a. Churn Category – Total Churn.
    b. Tooltip : Churn Reason – Total Churn.
6. Service Used:
    a. Internet Type – Churn Rate.
    b. prod_Service >> Services – Status – % RT Sum of Churn Status.

3.4 Predict Customer Churn (Using Python):

Create Churn Prediction Model – Random Forest:

- Importing Libraries & Data Load
- Data Preprocessing.
- Train Random Forest Model.
- Evaluate Model.
- Save Model for Prediction on New Predicton Data as CSV.

3.5 Power BI Visualization of Predicted Data:

- Import New Prediction CSV Data which created In Machine Learning model.
- Create Measures:
    1. Count Predicted Churner
    2. Title Predicted Churners
- Churn Prediction Page (Using New Predicted Data)

1. Right Side Grid:
    a. Customer ID
    b. Monthly Charge
    c. Total Revenue
    d. Total Refunds
    e. Number of Referrals
2. Demographic
    a. Gender – Churn Count
    b. Age Group – Churn Count
    c. Marital Status – Churn Count
3. Account Info
    a. Payment Method – Churn Count
    b. Contract – Churn Count
    c. Tenure Group - Churn Count
4. Geographic
    a. State – Churn Count

4 Resultat och Diskussion:

In this part, I will show some results which we find it and discussed it.

4.1 Results from Exploratory Data Analysis (EDA)

4.1.1 Finding Keys:

Demographic Insights:

- Total Customers: 6,418
- Total Stayed: 4,275
- Total Churned Customers: 1,732
- New Joiners: 411
- Gender: 63.07% of customers are female, and 36.93% are male.
- Contract Type: Most customers (90.2%) are on month-to-month contracts.
- Churn Rate: 27%.
- Total States: 22
- Total Revenue: 19.47 M
- Prediction Churn: 373 of 411 new customers which is 90%.
- Top 5 Churned Reason:
    1. Competitors have better offer.
    2. Competitors have better device.
    3. Attitude of support person.
    4. Price is too high.
    5. Product dissatisfaction.

4.2 Results from the Modelling Phase:

- ### 4.2.1 Random Forest:
  - Accuracy: 82%.
  - Precision: 83%.
  - Recall: 80%.
  - F1-score: 81%.

- ### 4.2.2 Logistik Regression:
  - Accuracy: 79%.
  - Precision: 81%.
  - Recall: 76%
  - F1-score: 78%

- ### 4.2.3 Decision Trees:
  - Accuracy: 74%
  - Precision: 72%
  - Recall: 75%
  - F1-score: 73%

- ### 4.2.4 Support Vector Machine (SVM):
  - Accuracy: 77%
  - Precision: 78%
  - Recall: 75%
  - F1-score: 76%

- ### 4.2.4 Support Vector Machine (SVM):
  - Accuracy: 80%
  - Precision: 82%
  - Recall: 77%
  - F1-score: 79%

5 Slutsatser and Diskussion

Churn analysis is a crucial process for businesses that want to reduce churn and retain customers. By understanding why customers churn, businesses can develop targeted retention strategies that address the specific needs of each customer segment.

Some of the most important conclusions from the work are:

- Critical factors for churn: most of churned are female, it is 65 % of total churned and most of them are more than 50 years old. To the company should concentrate to this customer group and give them better offers.
- Churn Reason: two important factors of churn is Customer service and competitor had better offer or devices. So , we should improve call center or customer service attitude , at the same time we should offer to customers cheap and better services.
- Total Churned by contract: we notice that, most churned customers (95 %) have month to month contract. In this case we should to review this contract and improve it.
- Using best model for prediction: Random forest can be computationally expensive, particularly when working with large datasets. It requires a lot of memory, which can be a constraint when working with limited resources. Although random forest is resistant to overfitting, it can still occur in certain cases, particularly when working with noisy data.

6 Självutvärdering

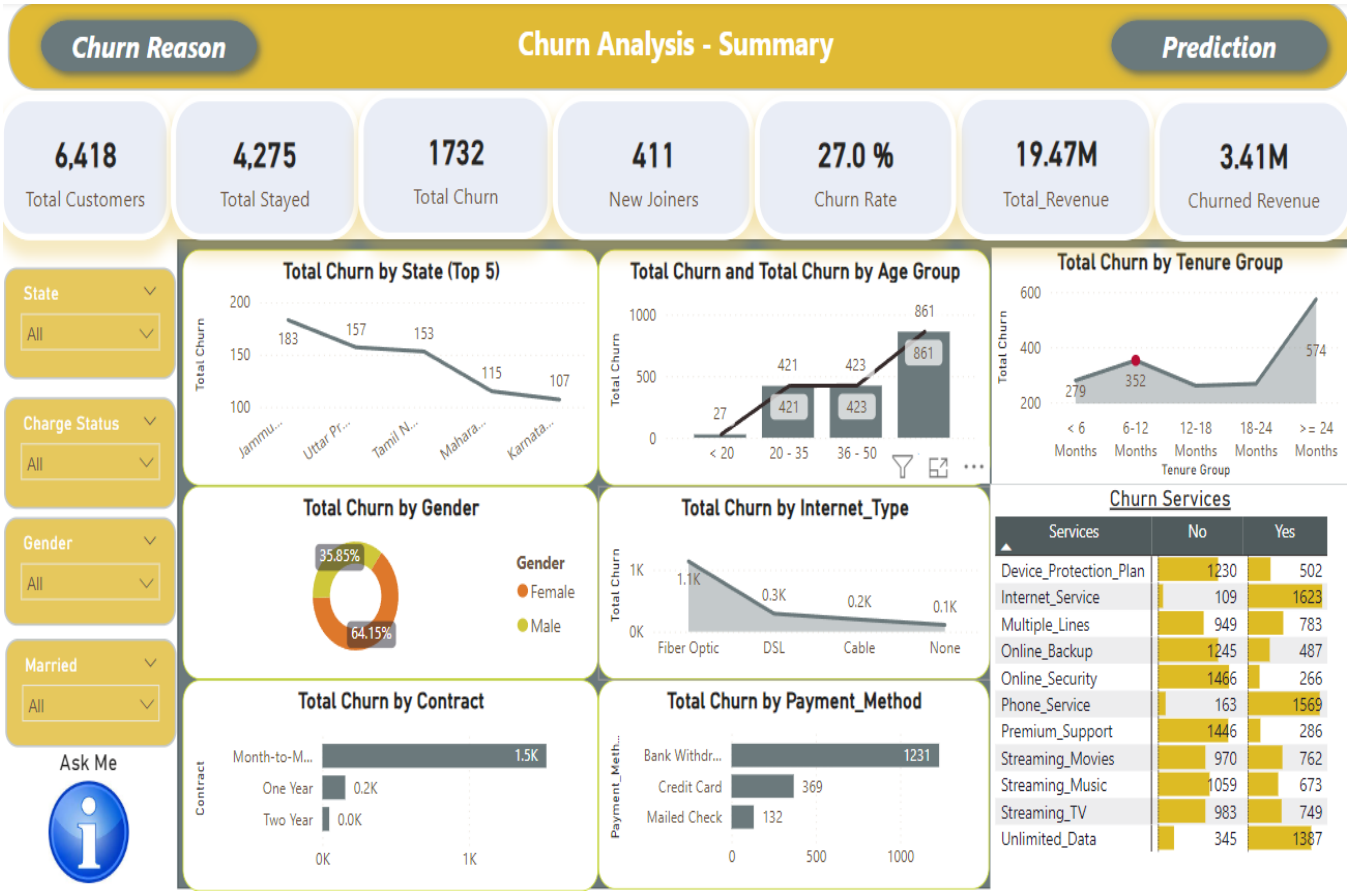
1. Utmaningar du haft under arbetet samt hur du hanterat dem.
    a. How to work as group.
    b. Loading data in SQL, PowerBI and Python Data
    c. Preparation Data (Removes nulls and checked datatype).
    d. Use different models for prediction.
    e. How to link data between SQL, Python and PowerBI.

2. Vilket betyg du anser att du skall ha och varför.
    a. The grade that represents my personal understanding of the material so I could better myself and apply my current knowledge to reach a higher comprehension.

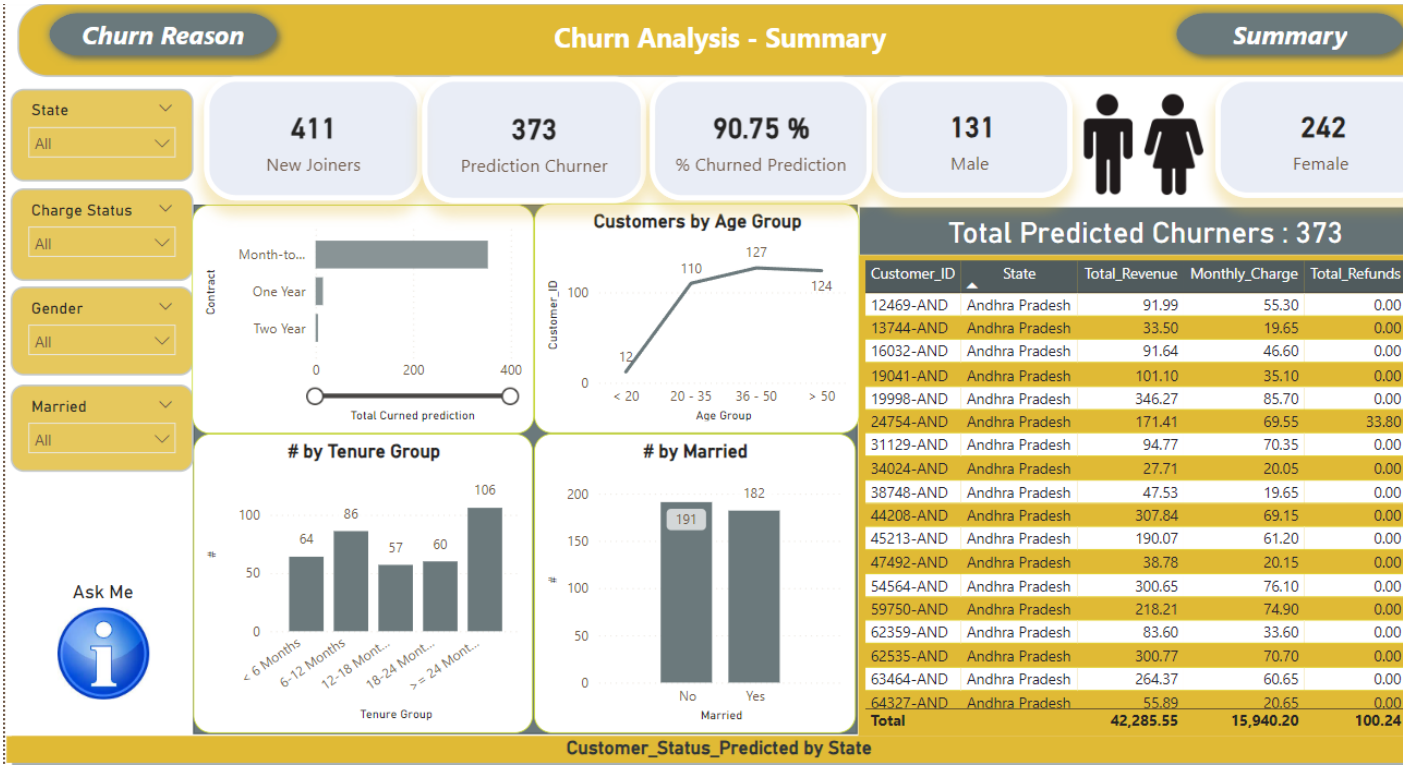
3. Något du vill lyfta fram till Antonio?
    a. I would like to thank you for your explaining, patient, helping, solving the problems and your support to all.

Appendix A

Summary



Prediktion:



Källförteckning

- Oracle Corporation. MySQL Documentation MySQL Tutorial - Learn MySQL Fast, Easy and Fun.
- SQL Tutorial for Beginners
- Microsoft Power BI documentation: Rutten ekonomi på jobbet. Inkomstförsäkring ingår.
- Python Software Foundation. Python. Welcome to Python.org
- Churn Analysis: How To Use AI To Analyze And Reduce Churn