# Project Proposal

On

# Customer Churn Prediction Using Python and Data Visualization

Guided By :-

Anuj Kumar

Created By: -

- Utkarsh

AFId : -

- AF04991819

Batch Code: ANP-D2868

Course Code: DAVIS-1

# Table of Contents


1. Title of the Project
2. Introduction
3. Objectives
4. Project Category
5. Analysis
    - Modules & Description
    - Database Design
    - ER Diagram
    - Data Flow Diagram
6. Complete Structure
    - Process Logical Diagram
7. Platform Used
    - Hardware Requirement
    - Software Requirement
8. Future Scope
9. Bibliography

# 1. Title of the Project

# Customer Churn Prediction Using Python and Data Visualization

# 2. Introduction

Customer churn refers to the situation where customers discontinue using a company's services over a certain period. Understanding why customers churn and identifying patterns behind their behavior is extremely important for industries such as telecommunications, banking, insurance, and subscription-based services. Retaining existing customers is far more cost-effective than acquiring new ones, making churn analysis a critical business activity.

This project, **"Customer Churn Analysis and Visualization Using Python,"** focuses on exploring customer-related data, identifying major trends, and visualizing the factors contributing to churn. The analysis is performed entirely using Python libraries such as Pandas, NumPy, Matplotlib, and Seaborn. Through data cleaning, preprocessing, and detailed visual exploration, the project uncovers insights related to customer tenure, service usage, contract type, monthly charges, payment modes, and demographic information.

The goal of this project is not to build a predictive model, but to **understand churn behavior through data visualization**, highlight key business insights, and help organizations design effective customer retention strategies. This analytical approach supports decision-making, reduces churn rate, and enhances overall customer satisfaction.

# 3. Objectives

1. **To analyze customer-related data** and understand patterns associated with customer retention and churn using Python.
2. **To identify key factors influencing churn**, such as contract type, service usage, tenure, monthly charges, and customer demographics.
3. **To perform detailed exploratory data analysis (EDA)** using Pandas for data handling and Matplotlib/Seaborn for visualization.
4. **To generate clear and meaningful visual representations**—including bar charts, histograms, heatmaps, and correlation maps—highlighting relationships between customer features and churn behavior.
5. **To design a structured analytical workflow** that includes data cleaning, feature exploration, and insight generation.
6. **To support decision-making for customer retention strategies** by providing actionable insights derived from the analysis.
7. **To prepare a comprehensive documentation report** covering analysis, diagrams, platform requirements, and future scope as per academic guidelines.

# 4. Project Category

**Data Analysis and Visualization Project**

This project falls under the category of Data Analytics, focusing specifically on Exploratory Data Analysis (EDA) and Data Visualization using Python programming. It involves collecting, cleaning, processing, analyzing, and visually interpreting customer churn data to extract meaningful business insights. No machine learning or predictive modeling techniques are used; instead, the project emphasizes understanding customer behavior through graphical and statistical analysis.

# 5. Analysis

The analysis phase focuses on understanding the project requirements, exploring the dataset, identifying system modules, designing workflows, and establishing the structure required to conduct customer churn analysis using Python. This section describes the functional flow of the project, data behavior, and analytical components that guide the development of the system.

---

### 5.1 Problem Definition

Customer churn occurs when customers discontinue their relationship with a company. The challenge for businesses is not only identifying customers who have churned but also understanding the reasons behind their exit.

This project analyzes the Telco Customer Churn dataset (or any similar churn dataset) to uncover:

- What characteristics differentiate churned customers from retained customers.
- Which customer attributes strongly correlate with churn.
- How demographic, service-related, and billing-related variables impact customer retention.

The project does not involve prediction. Instead, it focuses on insight generation through visualization.

---

### 5.2 Requirement Analysis

#### 5.2.1 Functional Requirements

1. Import and preprocess raw customer data using Python.
2. Perform exploratory data analysis (EDA) to understand patterns and distributions.
3. Visualize categorical and numerical attributes to compare churn vs non-churn behavior.
4. Generate correlation maps to identify relationships between variables.
5. Highlight the most influential factors contributing to customer churn.
6. Produce a detailed documentation/report summarizing analysis findings.

#### 5.2.2 Non-Functional Requirements

- Usability: Code must be simple, readable, and modular.
- Accuracy: Visuals should accurately represent real data patterns.

- Performance: Python operations should run efficiently on standard hardware.
- Portability: The project must run on any system with Python installed.
- Scalability: Can handle moderate-sized CSV datasets (5k–50k rows).

---

**5.3 Dataset Description**

The churn dataset typically contains features such as:

- Customer Demographics: gender, senior citizen, partner, dependents
- Account Information: tenure, contract type, payment method
- Services Used: internet service, phone service, online security, streaming
- Financial Features: monthly charges, total charges
- Target Attribute: churn (Yes/No)

These attributes allow multi-dimensional analysis of customer behavior.

---

**5.4 System Modules**

**Module 1: Data Collection & Loading**

- Load dataset (.csv) using Pandas
- Inspect structure (info, head, shape)
- Identify missing or inconsistent values

**Module 2: Data Cleaning & Preparation**

- Handle missing values
- Convert data types (e.g., TotalCharges to numeric)
- Filter inconsistent records
- Encode or restructure columns for analysis

**Module 3: Exploratory Data Analysis**

- Descriptive statistics (mean, median, frequency counts)
- Distribution analysis (histograms, boxplots)
- Comparison of churn vs non-churn groups

**Module 4: Data Visualization**

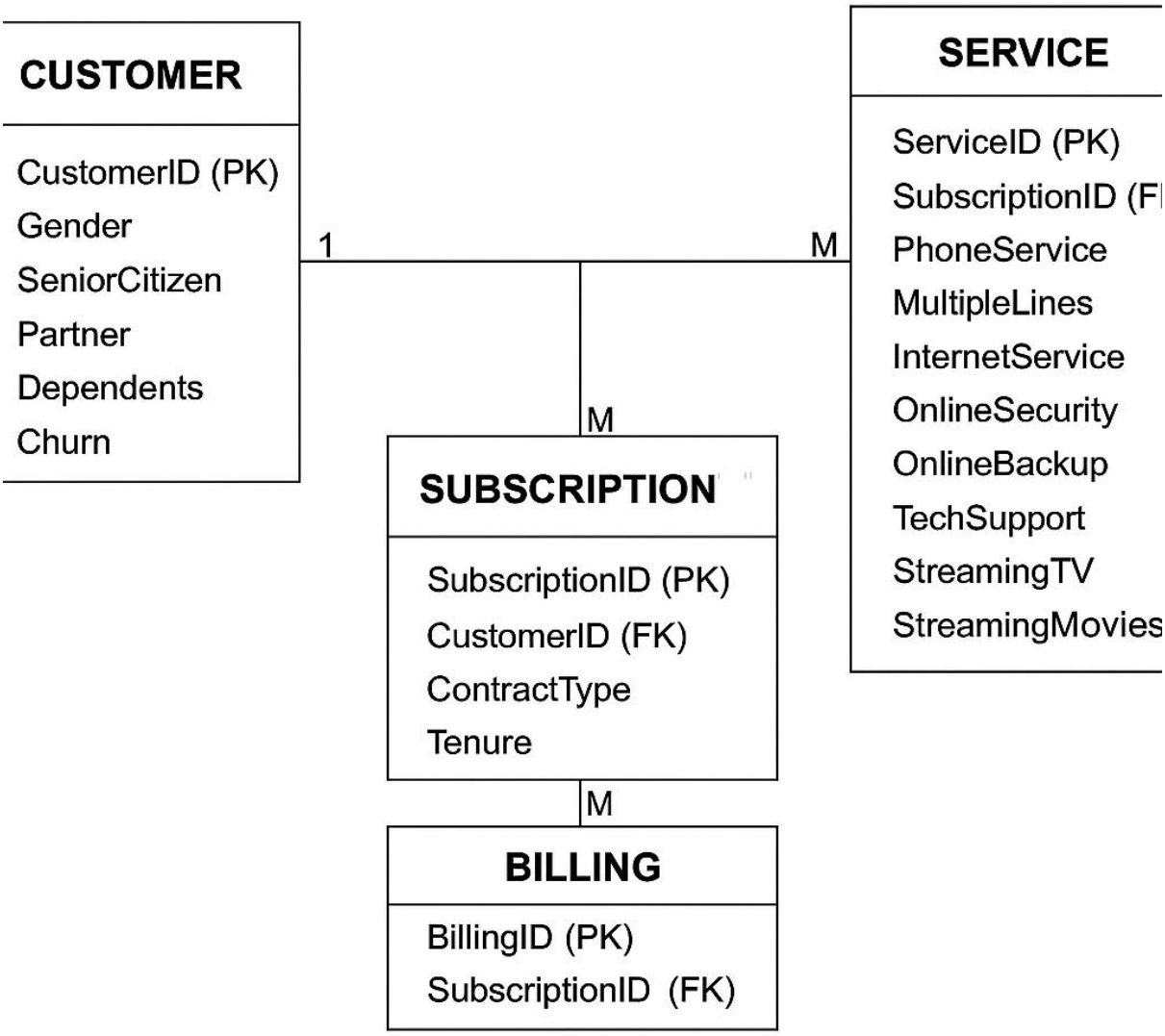
- Bar charts for categorical variables
- Heatmaps for correlation

- Pie charts for distribution percentages
- Pairwise comparisons (seaborn pairplots / countplots)
- Churn factor comparison (Contract type, Tenure, Charges)

**Module 5: Insight Extraction & Reporting**

- Highlight key findings
- Identify patterns linked to churn
- Recommend strategies (e.g., focus on low-tenure or high-charge customers)

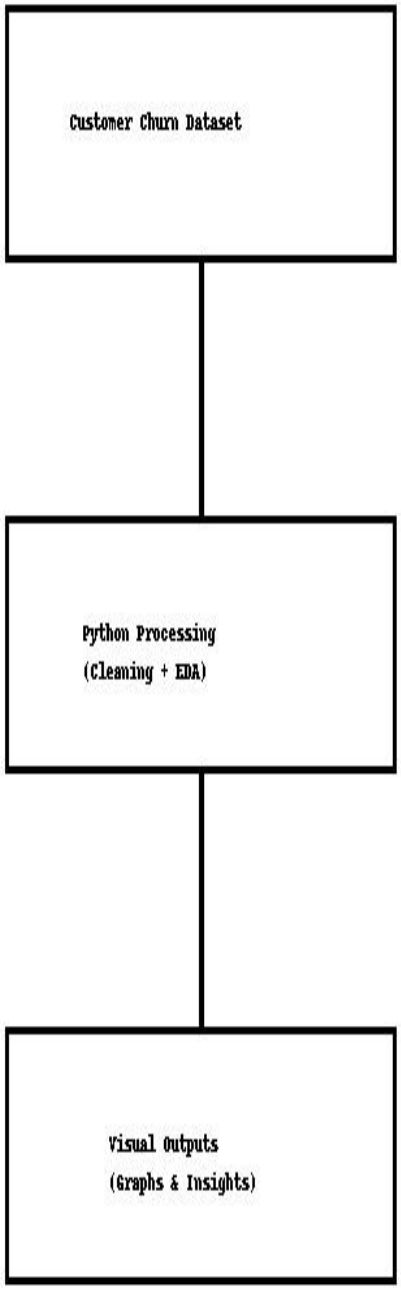
**ER Diagram**


CUSTOMER
CustomerID (PK)
Gender
SeniorCitizen
Partner
Dependents
Churn
1
M
SERVICE
ServiceID (PK)
SubscriptionID (F
PhoneService
MultipleLines
InternetService
OnlineSecurity
OnlineBackup
TechSupport
StreamingTV
StreamingMovies
M
SUBSCRIPTION
SubscriptionID (PK)
CustomerID (FK)
ContractType
Tenure
M
BILLING
BillingID (PK)
SubscriptionID (FK)

## Data Flow Diagram (DFD)

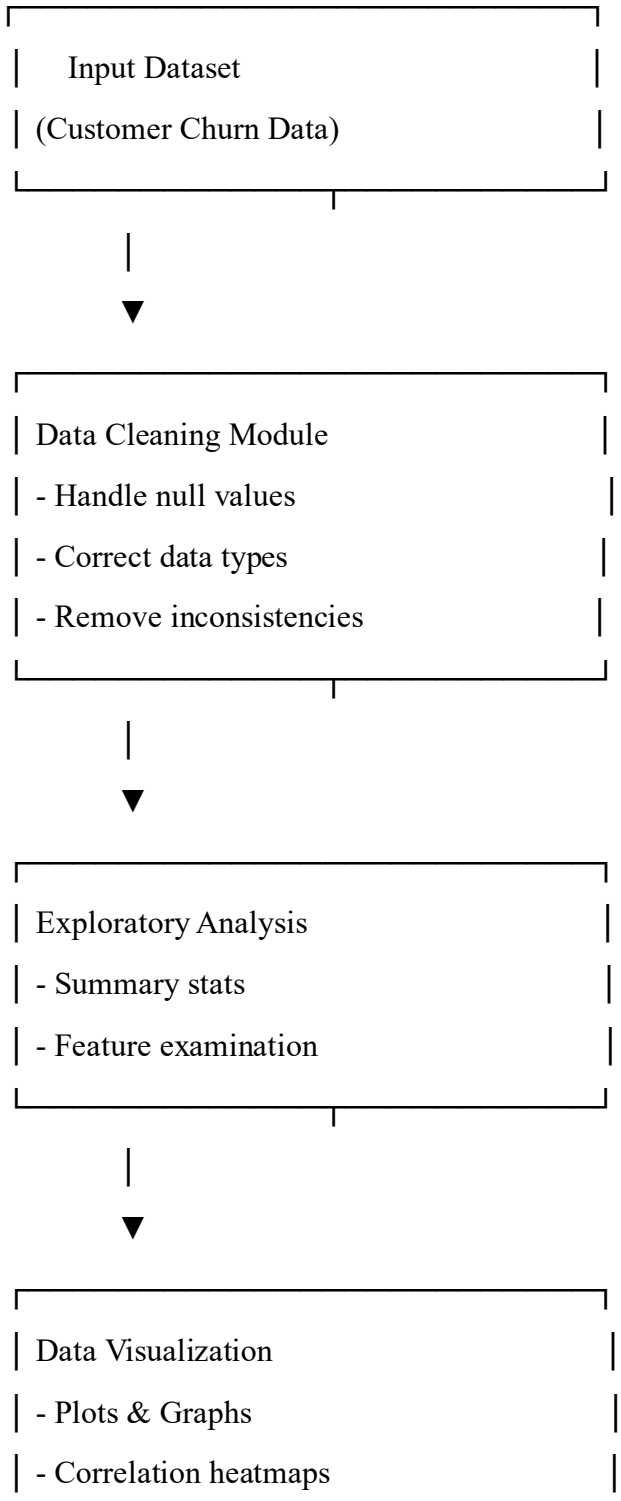
The Data Flow Diagram represents how data moves through the system from input to output.

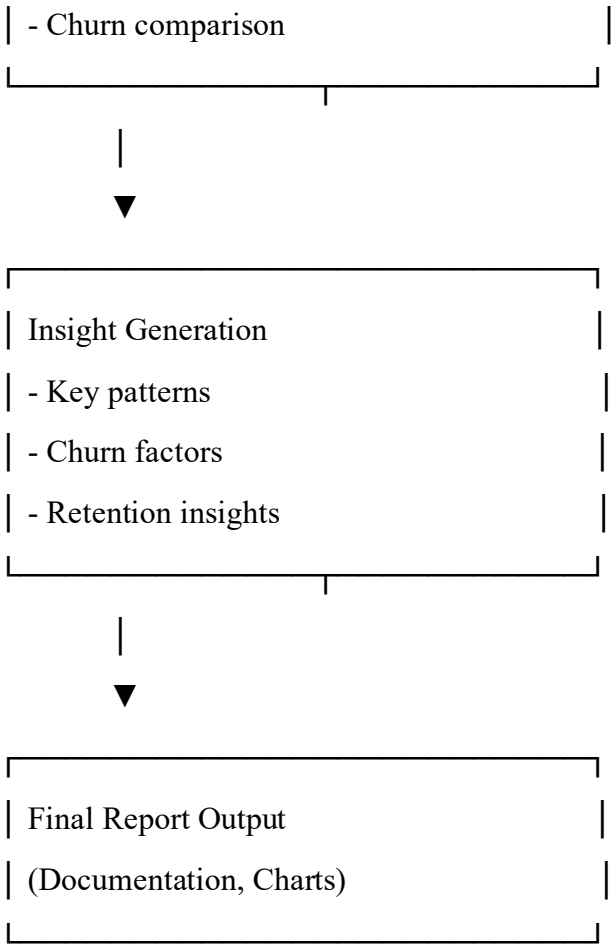
**DFD –**



# 6. Complete Structure

**Process Logical Diagram (Workflow)**




- Churn comparison
Insight Generation
- Key patterns
- Churn factors
- Retention insights
Final Report Output
(Documentation, Charts)

# 7. Platform Used

### 7.1 Hardware Requirements

| Component | Minimum Requirement | Recommended |
|---|---|---|
| Processor | Intel Core i3 / AMD equivalent | Intel Core i5 or higher |
| RAM | 4 GB | 8 GB or more |
| Storage | 1 GB free space | 2 GB free space |
| Graphics | Integrated graphics | Any basic graphics support |
| System Type | 32-bit / 64-bit | 64-bit preferred |

### 7.2 Software Requirements

| Software | Purpose |
|---|---|
| Operating System | Windows / Linux / macOS |
| Python (Version 3.7 or above) | Core programming language |
| Jupyter Notebook / VS Code / PyCharm | Development environment |
| **Python Libraries:** | |
| – Pandas | Data loading & manipulation |
| – NumPy | Numerical operations |
| – Matplotlib | Data visualization |
| – Seaborn | Statistical plots & heatmaps |
| – CSV Module | For dataset handling |
| Browser | To view documentation, plots (if exported), or references |

# 8. Future Scope

The current project focuses on customer churn analysis through Python-based data visualization. It can be extended in several ways in the future:

1. **Include Machine Learning Models** to predict customer churn instead of only analyzing it.
2. **Develop an Interactive Dashboard** using Streamlit or Dash for real-time, user-friendly analysis.
3. **Automate the Data Pipeline** so new data is cleaned, processed, and visualized automatically.
4. **Enable Real-Time Analytics** by integrating live data sources or APIs.
5. **Add Advanced Statistical Methods** for deeper insight into customer behavior.
6. **Include Customer Segmentation** to identify different customer groups and tailor retention strategies.
7. **Integration with Business Tools** like CRM systems for practical organizational use.

# 9. Bibliography


1. Wes McKinney, *Python for Data Analysis*, O'Reilly Media, 2018.
2. Jake VanderPlas, *Python Data Science Handbook*, O'Reilly Media, 2016.
3. Pandas Documentation – https://pandas.pydata.org
4. NumPy Documentation – https://numpy.org
5. Matplotlib Documentation – https://matplotlib.org
6. Seaborn Documentation – https://seaborn.pydata.org
7. IBM Sample Data: Telco Customer Churn Dataset (commonly used for churn analysis).
8. TutorialsPoint – Python Data Science and Visualization Resources.
9. GeeksforGeeks – Python Data Analysis & Visualization Articles.
10. Kaggle – Customer Churn Datasets and Data Exploration Notebooks.