Customer Churn Prediction Using Python and Data Visualization

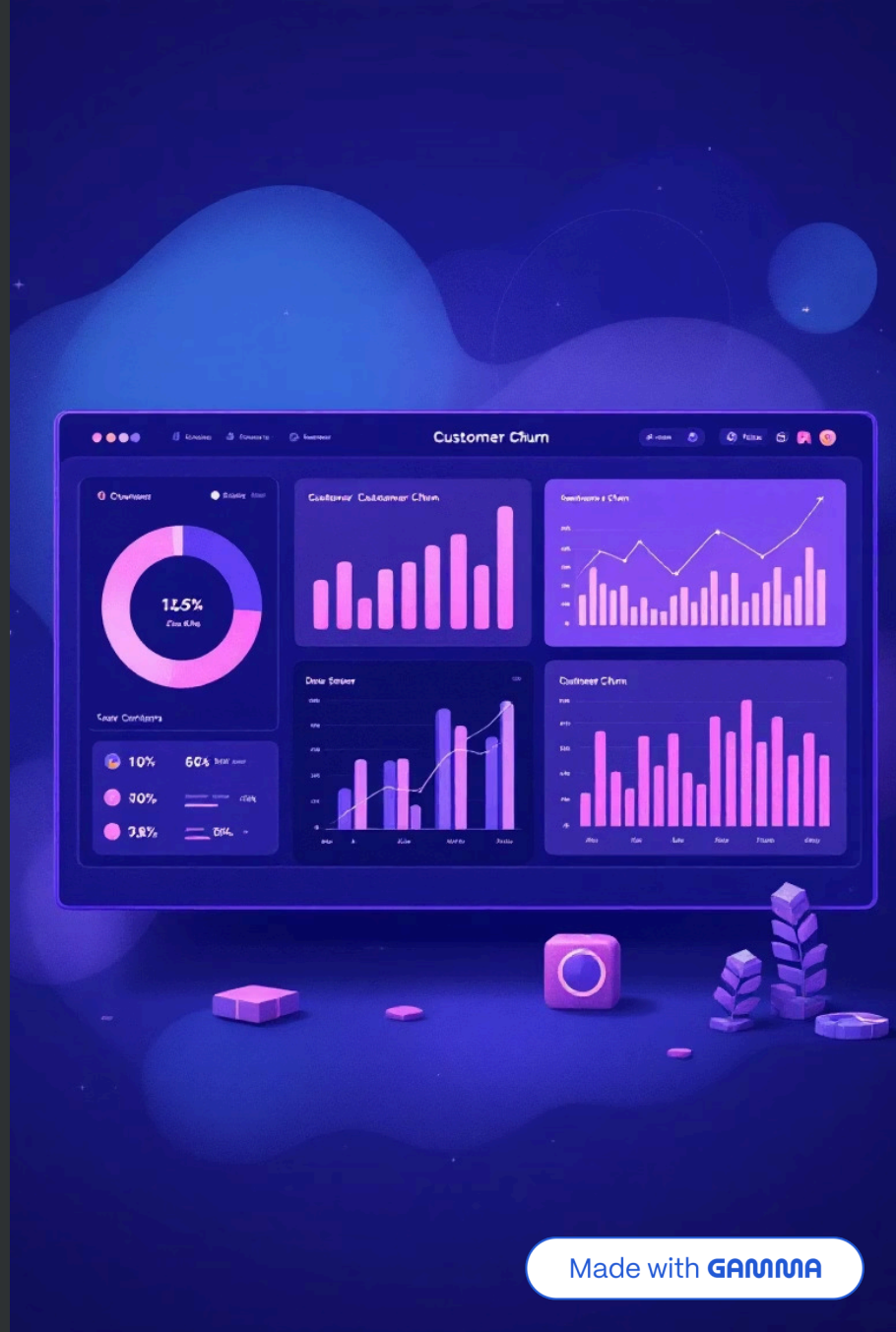
A project proposal by Utkarsh (AF04991819), Batch ANP-D2868, Course DAVIS-1.

Introduction to Customer Churn

Customer churn occurs when clients stop using a company's services. Retaining existing customers is more cost-effective than acquiring new ones, making churn analysis vital for industries like telecom, banking, and subscription services.

Project Overview: Churn Analysis & Visualization

This project, "Customer Churn Analysis and Visualization Using Python," explores customer data to identify trends and visualize factors contributing to churn. It uses Python libraries like Pandas, NumPy, Matplotlib, and Seaborn for data cleaning, preprocessing, and detailed visual exploration.



Key Objectives

1

Analyze Data

Understand customer retention and churn patterns.

2

Identify Factors

Pinpoint influences like contract type, service usage, and demographics.

3

Perform EDA

Detailed exploratory data analysis using Python libraries.

4

Generate Visuals

Create clear charts and maps for relationships.

5

Support Decisions

Provide actionable insights for retention strategies.

Project Category: Data Analysis & Visualization

This project focuses on Exploratory Data Analysis (EDA) and Data Visualization using Python. It involves collecting, cleaning, processing, analyzing, and visually interpreting customer churn data to extract meaningful business insights, without machine learning or predictive modeling.

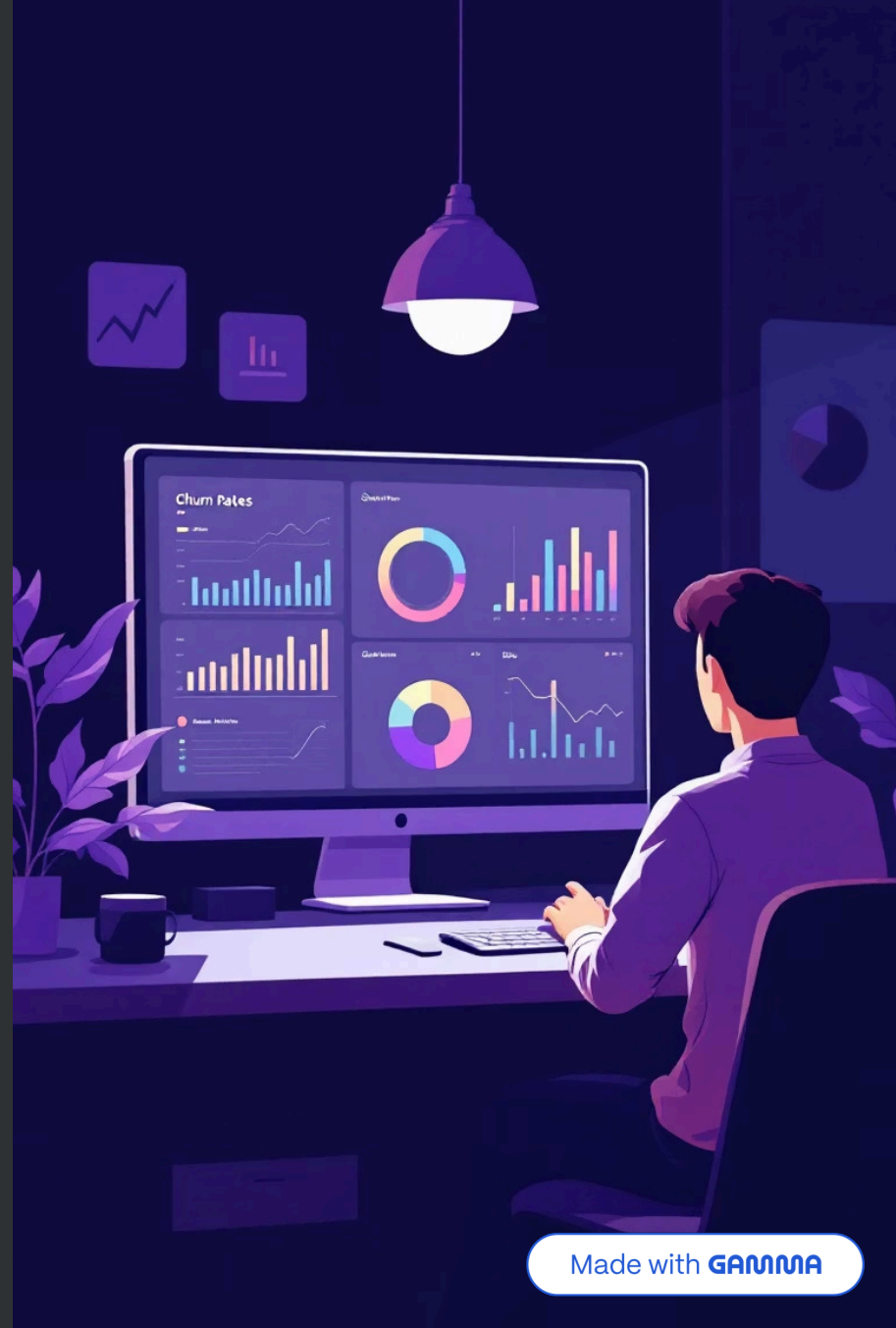
Analysis: Problem & Requirements

Problem Definition

Identify characteristics and attributes that differentiate churned from retained customers, and how demographics, service, and billing impact retention. Focus on insight generation through visualization.

Functional Requirements

- Import & preprocess data.
- Perform EDA & visualize attributes.
- Generate correlation maps.
- Highlight influential churn factors.
- Produce detailed documentation.



System Modules & Data Flow

1

Data Collection & Loading

Load CSV, inspect structure, identify missing values.

2

Data Cleaning & Preparation

Handle missing values, convert types, filter records.

3

Exploratory Data Analysis

Descriptive statistics, distribution analysis, churn comparison.

4

Data Visualization

Bar charts, heatmaps, pie charts, pairwise comparisons.

5

Insight Extraction & Reporting

Highlight findings, identify patterns, recommend strategies.

Platform Requirements

Hardware

Processor	Intel Core i5+
RAM	8 GB or more
Storage	2 GB free space
System Type	64-bit preferred

Software

- Operating System: Windows / Linux / macOS
- Python (3.7+)
- Jupyter Notebook / VS Code
- Libraries: Pandas, NumPy, Matplotlib, Seaborn

Future Scope & Expansion

Machine Learning Models

Predict customer churn.

Interactive Dashboard

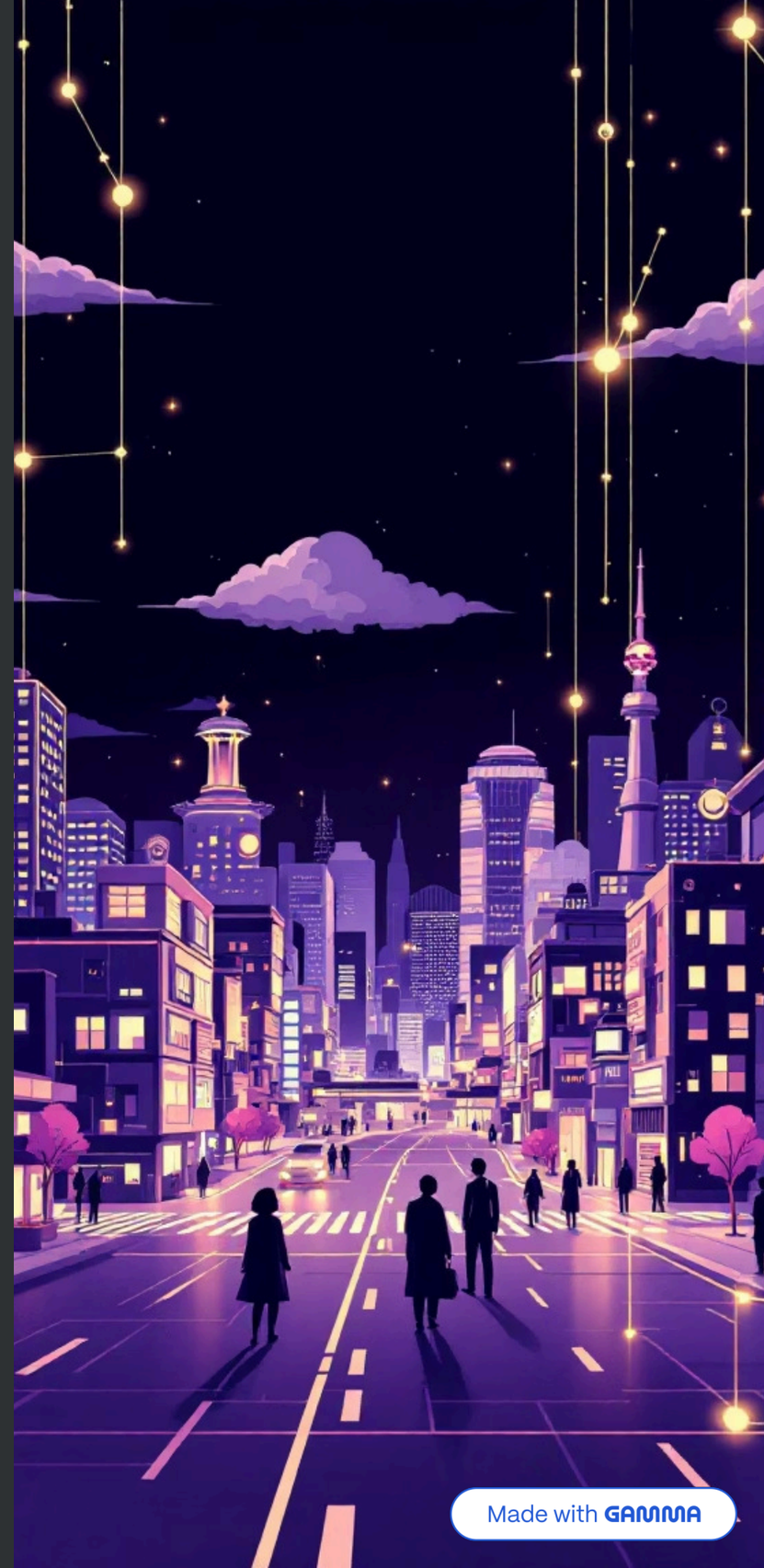
Real-time, user-friendly analysis.

Automate Data Pipeline

Automatic cleaning, processing, visualization.

Customer Segmentation

Tailor retention strategies.



Bibliography

- Wes McKinney, *Python for Data Analysis*, O'Reilly Media, 2018.
- Jake VanderPlas, *Python Data Science Handbook*, O'Reilly Media, 2016.
- Pandas Documentation – **https://pandas.pydata.org**
- NumPy Documentation – **https://numpy.org**
- Matplotlib Documentation – **https://matplotlib.org**
- Seaborn Documentation – **https://seaborn.pydata.org**
- IBM Sample Data: Telco Customer Churn Dataset.
- TutorialsPoint – Python Data Science and Visualization Resources.
- GeeksforGeeks – Python Data Analysis & Visualization Articles.
- Kaggle – Customer Churn Datasets and Data Exploration Notebooks.