Predicting Term Deposit Subscription in Bank Marketing Using Big Data Analytics

Student: Jospin Nabonyimana (ID: 26511)

Course: INSY 8413 – Introduction to Big Data Analytics

Institution: Adventist University of Central Africa

Date: Monday, August 4th, 2025

Project Introduction & Goals

Background

- Banks face low response rates in marketing campaigns.
- Data-driven targeting can significantly improve success.
- Inefficient outreach wastes valuable resources.

Project Goal

- Analyze client and campaign data using Python & Power BI.
- Predict which clients are most likely to subscribe to a term deposit.
- Enhance strategic decision-making in marketing.

Dataset Overview

Source: UCI Bank Marketing Dataset

Size: 45,211 rows, 17 columns

Nature: Tabular, structured CSV format

Key Variables: Age, job, education, balance, duration, campaign, etc.

Target Variable: 'y' (client subscribed to a term deposit: Yes/No)

Methodology & Approach

1 Data Cleaning

Replaced 'unknown' values with nulls and performed imputation. Addressed outliers, specifically focusing on zero duration calls and extreme balance figures to ensure data integrity.

2 Data Transformation

Applied one-hot encoding for categorical variables. Converted relevant fields to a numeric format to prepare data for model training.

3 Exploratory Data Analysis (EDA)

Utilized Seaborn and Matplotlib for comprehensive visualizations. Investigated relationships between job, education, balance, duration, and term deposit subscriptions.

4 Modeling

Implemented Logistic Regression with an 80/20 train/test split. Evaluated model performance using Accuracy, Confusion Matrix, and Classification Report for robust assessment.

Results & Key Visualizations

Model Performance

Accuracy: 83.7%

Precision for 'Yes': 69%

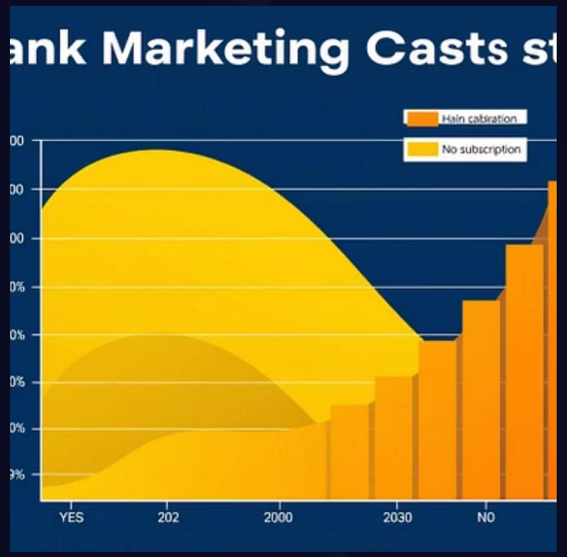
- The model successfully identified key features influencing subscription rates, providing actionable insights for marketing optimization.

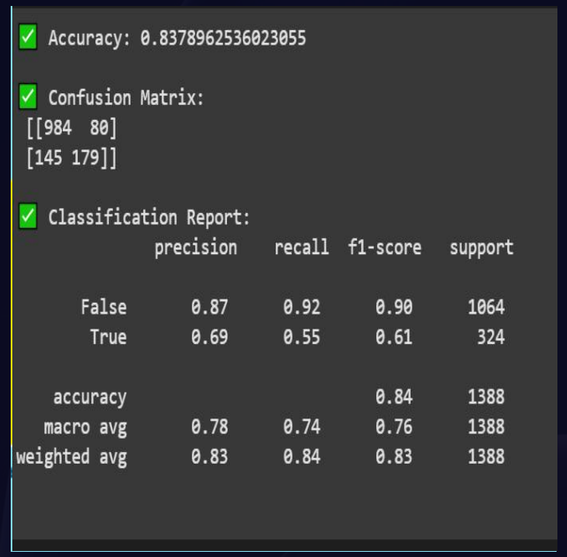
Key Graphs & Insights

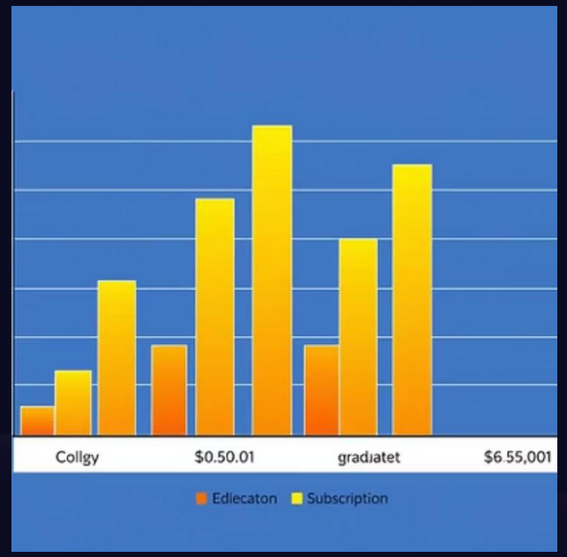
Countplot: Subscription vs. Non-Subscription distribution.

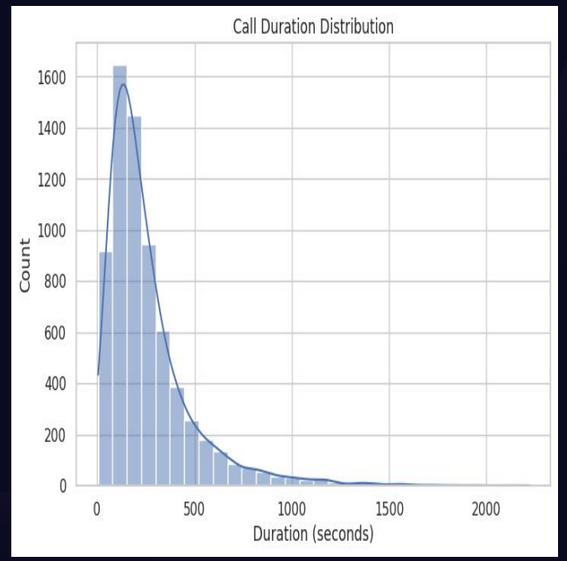
Bar Charts: Job vs. Subscription, Education vs. Subscription. Retired and management clients show higher subscription rates; education level correlates with outcome.

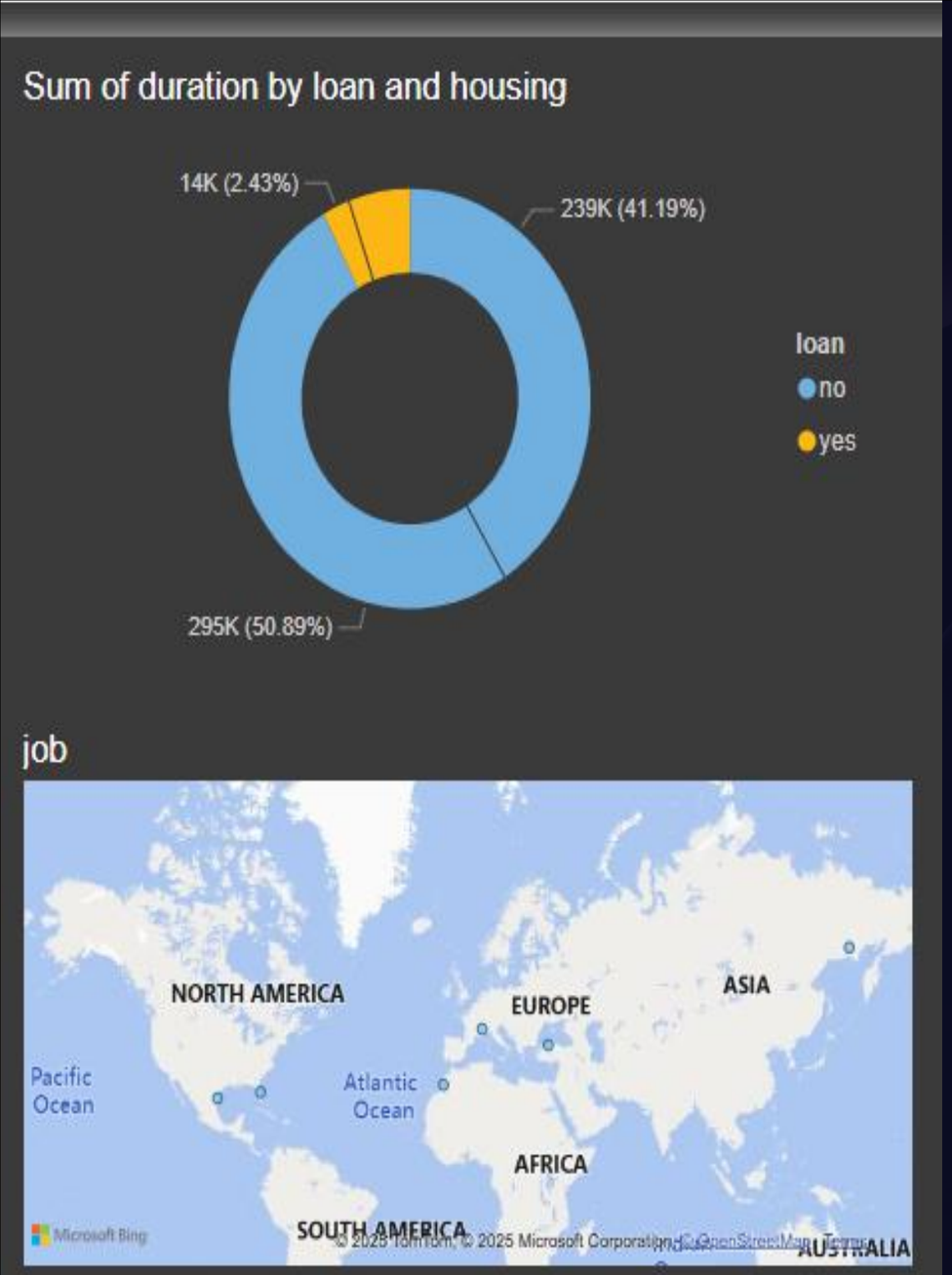
Histogram: Call Duration Distribution. Longer call durations significantly lead to more subscriptions, highlighting the importance of engaged conversations.











Power BI Dashboard Overview

The interactive Power BI dashboard provides a dynamic interface for exploring the bank marketing data.

Dashboard Features

- Imported cleaned data from Python processing.
- Interactive slicers for Education Level, Job Role, and Subscription Status, enabling dynamic filtering.
- Filters for visuals: Education, Job, Month, Contact Method.

Dashboard Visuals

Subscription Distribution: Pie Chart

Job vs. Subscription: Stacked Bar Chart

Education vs. Subscription: Bar Chart

Call Duration: Histogram

Subscription by Month: Line Chart

Strategic Recommendations

Based on our comprehensive analysis, we provide data-driven recommendations to optimize bank marketing campaigns.

Prioritize Call Duration

Focus marketing efforts on clients with higher engagement during calls, as longer call durations strongly correlate with increased subscription rates.

Target Key Demographics

Direct campaigns towards individuals in "Retired" and "Management" roles, who consistently demonstrate higher propensities for term deposit subscriptions.

Segment by Education

Implement refined segmentation strategies based on education levels to tailor messaging and offers, leading to more effective targeting and improved conversion.

Avoid Low-Probability Segments

Minimize resource allocation to segments with historically low subscription rates to prevent wasted effort and maximize return on investment.

Future Work & Enhancements

Our project lays a strong foundation for future analytical expansions.

Implement Advanced Models

Explore and integrate more sophisticated machine learning algorithms such as Random Forest, Gradient Boosting (XGBoost), or Support Vector Machines for potentially higher predictive accuracy.

Address Class Imbalance

Utilize techniques like SMOTE (Synthetic Minority Over-sampling Technique) to balance the target variable classes, improving the model's ability to predict the minority class (subscribers).

Expand Dataset

Augment the current dataset with additional demographic, socioeconomic, or macroeconomic information to uncover deeper insights and enhance predictive power.

Enhance Power BI Interactivity

Incorporate advanced Key Performance Indicators (KPIs), custom visuals, and more interactive elements within the Power BI dashboard to facilitate dynamic decision-making and deeper exploration.

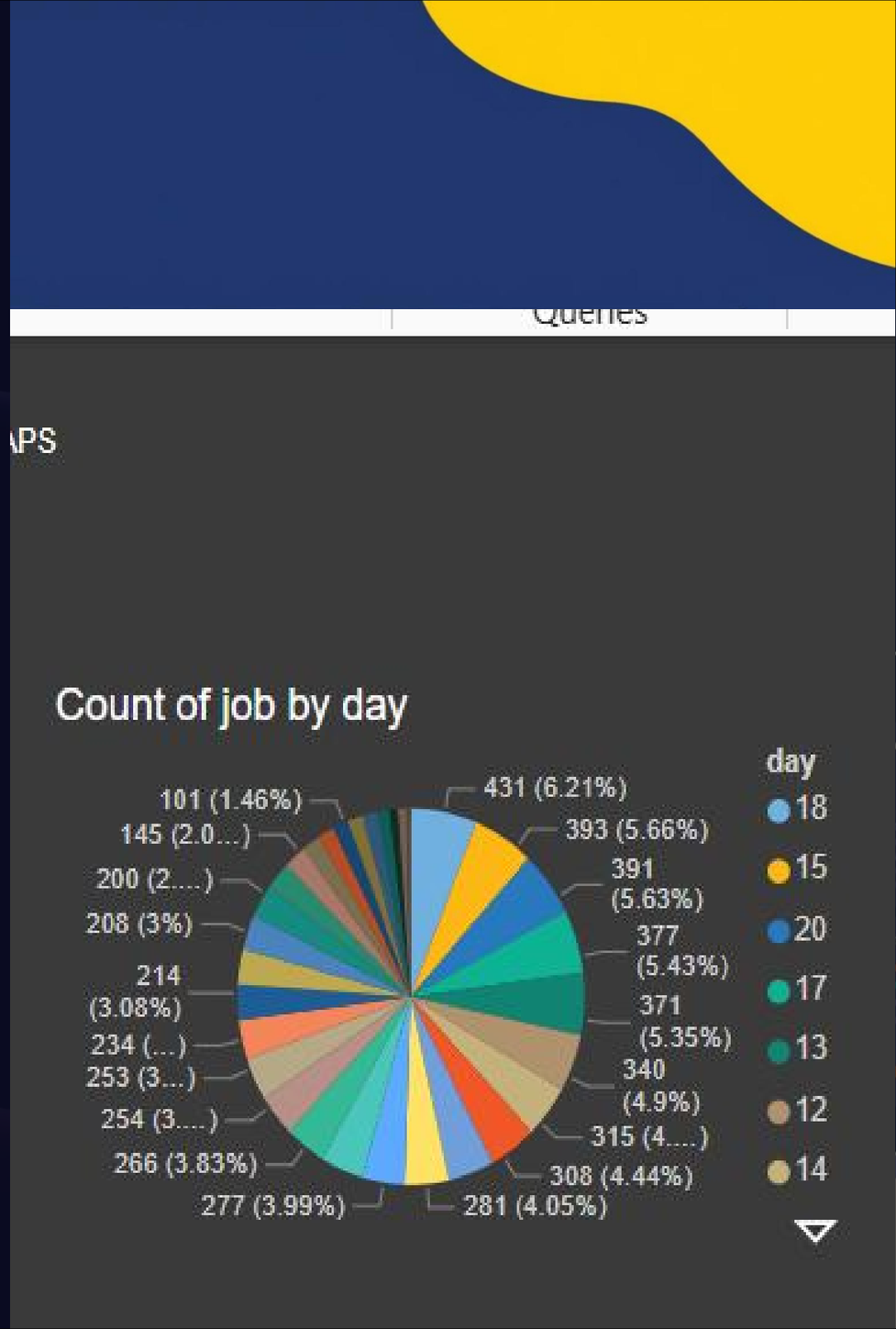
Submission Requirements

GitHub Repository

- Well-structured folders for code, data, and documentation.
- Comprehensive README.md file with project overview, setup instructions, and screenshots.
- All Python code files (.ipynb, .py) and the Power BI project file (.pbix) included.

PowerPoint Presentation

- A concise slide deck summarizing: Project Introduction, Methodology, Detailed Results, Strategic Recommendations, and Future Work.
- Visuals and key takeaways clearly presented to convey insights effectively.



Academic Integrity & Guidelines

This capstone project is strictly reflecting my original work and understanding. Adherence to these guidelines is paramount.

Original Work

The project's scope (title, sector, problem, dataset) must be submitted for prior verification(Via GitHub). I ensure that all aspects reflect your unique contribution.

No Plagiarism

Significant similarities in implementation, structural design, or code will lead to severe penalties, including low marks or a zero grade for the assignment, I didn't do.

I am ready to explain fully what I did in this project

Confidentiality

Do not share your project work (code, report, dashboard, or GitHub repository) with any individual outside of this specific course project. Maintain strict confidentiality.

YOU CAN GET MORE BY SIMPY HITTING THE FOLLOWING LINKS:

Github repo link: https://github.com/nabonyimanajospin/bank-marketing-subscription-analysis/blob/main/README.md
Google colab link: https://colab.research.google.com/drive/131Nciw5gYkaNowLrSGzRqqYVFGZ6UfQu
Dataset link: https://archive.ics.uci.edu/ml/datasets/Bank+Marketing

Conclusion and thanking

"Whatever you do, work at it with all your heart, as working for the Lord, not for human masters." — Colossians 3:23 (NIV)

Summary

- Successfully cleaned, analyzed, and modeled bank marketing data
- Discovered key insights to improve marketing campaigns
- Built a Power BI dashboard for decision-making
- Ready for future improvement and application

END OF MY PROJECT PRESENTATION

THANK YOU ALL