

Index:

1. Introduction and Project Objective
2. Data Collection and Preparation
3. Data Cleaning
4. Data Exploration and Analysis
5. Feature Engineering
6. Churn Prediction Model
7. SQL Database Design and Implementation
8. Visualizations and Reporting
9. Challenges and Learnings
10. Conclusion and Future Work
11. Introduction and Project Objective:
 - Objective: The objective of this project is to analyze customer retention and churn patterns for an e-commerce platform using data-driven insights. We aim to build a predictive model to identify potential churned customers and suggest strategic actions to improve retention.
12. Data Collection and Preparation:
 - Data Sources: Synthetic data was generated for three main datasets:
 - Customers Data
 - Transactions Data
 - Engagements Data
 - Libraries Used: pandas, numpy
 - Functions and Logic:

- Data loading using `pd.read_csv()`
- Basic data inspection using `df.head()`, `df.info()`, `df.describe()`

13. Data Cleaning:

- Objective: Handle missing values, correct data types, and standardize column names.
- Libraries Used: pandas, numpy
- Functions and Logic:
 - Column names standardized to lowercase using `df.columns = df.columns.str.lower()`
 - Missing values handled using `df.fillna()` and `df.dropna()`
 - Data types adjusted using `pd.to_datetime()` for dates, `astype()` for numerics.
 - Outlier detection using z-score method.

14. Data Exploration and Analysis:

- Objective: Analyze customer demographics, transaction patterns, and engagement levels.
- Libraries Used: pandas, matplotlib, seaborn, scipy, statsmodels
- Functions and Logic:
 - Descriptive statistics using `df.describe()`
 - Churn rate analysis using customer tenure and transaction frequency.
 - Time series analysis using ARIMA model to forecast sales trends.
 - Clustering using K-Means to segment customers based on age and transaction amount.

15. Feature Engineering:

- Objective: Extract relevant features to improve model accuracy.
- Libraries Used: pandas, numpy, sklearn
- Functions and Logic:
 - Creation of Recency, Frequency, and Monetary (RFM) metrics.
 - Binning customer ages into categorical segments.
 - Extraction of year, month, and day components from transaction dates.

16. Churn Prediction Model:

- Objective: Develop and evaluate a churn prediction model.
- Libraries Used: scikit-learn, xgboost, imbalanced-learn
- Functions and Logic:
 - Data split using `train_test_split()`
 - Model training using Logistic Regression, Random Forest, and XGBoost.
 - Evaluation using `accuracy_score()`, `precision_score()`, `recall_score()`, and ROC-AUC score.

17. SQL Database Design and Implementation:

- Objective: Store cleaned and processed data for structured querying.
- Database: SQLite
- Tables: customers, transactions, engagements
- SQL Scripts: Data insertion, querying, and analysis.

18. Visualizations and Reporting:

- Objective: Present key findings using visualizations.
- Libraries Used: matplotlib, seaborn

- Visualizations:
 - Churn rate trends over time.
 - Transaction frequency distribution.
 - Cluster analysis visualization.

19. Challenges and Learnings:

- Challenges encountered: Data inconsistencies, missing values, model tuning.
- Learnings: Importance of data preprocessing, feature selection, and hyperparameter tuning.

20. Conclusion and Future Work:

- Summary of findings and their implications.
- Recommendations for further analysis and model improvements.