

10-Day Plan for Building 'Customer Segmentation for Retail Business' Project

Day 1: Project Setup and Data Acquisition

Tasks:

- Define the project scope and objectives clearly.
- Identify and download the retail sales dataset.
- Set up your working environment, including SQL, Python, and Power BI.

Technologies: SQL, Python, Power BI.

Resources: Retail sales dataset, Python libraries (Pandas, NumPy).

Day 2: Data Exploration and Understanding

Tasks:

- Import the dataset into SQL and Python.
- Perform initial data exploration to understand the structure, types, and quality of the data.
- Generate summary statistics and visualizations to identify patterns and anomalies.

Technologies: SQL, Python (Pandas, Matplotlib/Seaborn).

Resources: Jupyter Notebook for Python exploration, SQL queries for data summarization.

Day 3: Data Cleaning and Preprocessing

Tasks:

- Handle missing values, duplicates, and any data inconsistencies.
- Standardize categorical variables and normalize numerical features if necessary.
- Split the data into training and testing sets for model validation later.

Technologies: SQL, Python (Pandas, Scikit-learn).

Resources: SQL scripts for data cleaning, Python scripts for preprocessing.

Day 4: Feature Engineering

Tasks:

- Create new features that could improve the clustering model, such as customer lifetime value, average purchase frequency, etc.
- Select the most relevant features for the clustering analysis.

Technologies: Python (Pandas, NumPy).

Resources: Python scripts for feature creation and selection.

Day 5: Clustering Model Selection and Implementation

Tasks:

- Research and select appropriate clustering algorithms (e.g., K-Means, Hierarchical Clustering).
- Implement the selected clustering algorithm on the preprocessed data.
- Determine the optimal number of clusters using methods like the Elbow method or Silhouette analysis.

Technologies: Python (Scikit-learn, Matplotlib).

Resources: Python scripts for model implementation, research papers/tutorials on clustering.

Day 6: Model Evaluation and Refinement

Tasks:

- Evaluate the clustering results using appropriate metrics (e.g., Silhouette score).
- Visualize the clusters to understand the segmentation better.
- Refine the model by adjusting parameters or features if needed.

Technologies: Python (Scikit-learn, Matplotlib/Seaborn).

Resources: Python scripts for model evaluation and visualization.

Day 7: Data Analysis and Insights Extraction

Tasks:

- Analyze the clusters to derive meaningful insights about customer behavior.
- Identify key characteristics of each customer segment.
- Document the findings with a focus on actionable business insights.

Technologies: Python, Power BI.

Resources: Python scripts for analysis, Power BI for visualizing the segments.

Day 8: Visualization in Power BI

Tasks:

- Import the segmentation results into Power BI.
- Create dashboards to visualize customer segments, highlighting key metrics and trends.
- Ensure the dashboards are interactive and provide a clear narrative.

Technologies: Power BI.

Resources: Power BI dataset (imported from Python/SQL), Power BI templates.

Day 9: Reporting and Documentation

Tasks:

- Compile the project into a final report, including methodology, findings, and visualizations.
- Create a presentation summarizing the project for your portfolio.

Technologies: Power BI, MS Word/Google Docs.

Resources: Power BI reports, documentation tools (MS Word, Google Docs).

Day 10: Review and Portfolio Integration

Tasks:

- Review the entire project, making any necessary refinements.
- Integrate the project into your portfolio, ensuring it is well-documented and presented.
- Prepare a summary or case study for your portfolio.

Technologies: Power BI, Portfolio platform (e.g., GitHub, personal website).

Resources: Portfolio hosting tools, documentation tools.