Title: Telco Customer Churn Analysis and Retention Strategy Development

1. Introduction

The objective of this project is to analyze a dataset from a Telecommunications company to identify the characteristics and behaviors of customers who have churned, i.e., left the company. By understanding these factors, predictive models can be developed to forecast customer churn, which in turn can inform strategies for customer retention.

2. Data Preparation

2.1 Importing Necessary Libraries

The following libraries are imported to facilitate data manipulation, visualization, and analysis:

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

2.2 Data Loading and Cleaning

2.2.1 Loading Dataset

The dataset is loaded into a DataFrame called `telco` from the specified CSV file path.

```
telco = pd.read_csv(r"path_to_file")
```

2.2.2 Inspecting Data

Initial inspection of the data is performed to understand its structure, dimensions, and data types, using methods like `head`, `tail`, `shape`, `dtypes`, and `nunique`.

```
telco.head()
telco.tail()
telco.shape
telco.dtypes
telco.nunique()
```

2.2.3 Handling Missing and Incorrect Values

The `TotalCharges` column had some rows with white spaces which are handled by dropping those rows and converting the column data type to float for further analysis.

```
telco = telco[telco['TotalCharges'] != ' ']
telco['TotalCharges'] = telco['TotalCharges'].apply(pd.to_numeric)
```

2.3 Data Exploration

2.3.1 Descriptive Statistics

Descriptive statistics are calculated to understand the central tendency, dispersion, and shape of the dataset’s distribution.

```
telco.describe()
```

2.3.2 Checking for Outliers

Outliers in the `TotalCharges` column are identified and handled to ensure they do not skew the analysis.

```
sns.boxplot(x= telco['TotalCharges'])
Q1 = telco.quantile(0.25)
Q3 = telco.quantile(0.75)
IQR = Q3 - Q1
telco_out = telco[~((telco < (Q1 - 1.5 * IQR)) | (telco > (Q3 + 1.5 *
IQR))).any(axis=1)]
```

3. Feature Engineering

Column contents are checked and data types are converted as necessary to prepare for model development. For example, categorical columns are one-hot encoded, and the `Churn` column is label encoded.

```
telco['TotalCharges'] = telco['TotalCharges'].apply(pd.to_numeric)
# Additional encoding and feature engineering code
```

4. Model Development

The data is now ready for developing predictive models to analyze customer churn.

```
# Model development code
```

5. Evaluation and Interpretation

The performance of the models is evaluated and the results are interpreted to understand the factors contributing to customer churn.

```
# Evaluation and interpretation code
```

6. Conclusion and Recommendations

Insights from the analysis are used to formulate recommendations for customer retention strategies.

```
# Conclusion and recommendations
```

This documentation provides an organized overview of the code and the steps undertaken to prepare the data for analysis, develop predictive models, and draw insights for developing customer retention strategies.