

```
/* Import the dataset */
proc import datafile="/home/u64091795/sasuser.v94/Airbnb_Analytics_Dataset.csv"
    out=airbnb_data
    dbms=csv
    replace;
    getnames=yes;
run;

/* Check for missing values in each column */
proc means data=airbnb_data n nmiss;
run;

/* Impute missing values with the mean */
proc stdize data=airbnb_data out=airbnb_imputed reponly missing=mean;
run;

/* Descriptive statistics */
proc means data=airbnb_imputed mean median std min max;
title "Descriptive Statistics";
run;
title;

/* Histogram of Stay Duration */
proc sgplot data=airbnb_imputed;
    histogram Stay_Duration;
    title "Histogram of Stay Duration";
run;
title;

/* Bar Chart of Property Type */
proc sgplot data=airbnb_imputed;
    vbar Property_Type;
    title "Bar Chart of Property Type";
run;
title;

/* Boxplot of Price per Night by Property Type */
proc sgplot data=airbnb_imputed;
    vbox Price_per_Night / category=Property_Type;
    title "Boxplot of Price per Night by Property Type";
run;
title;

/* Pie Chart for Property Type Distribution */
proc gchart data=airbnb_imputed;
    pie Property_Type / value=inside percent=outside;
    title "Pie Chart of Property Type Distribution";
run;
quit;
title;

/* Generate Correlation Matrix as a Dataset for Heatmap */
proc corr data=airbnb_imputed outp=corr_matrix nosimple;
    var Stay_Duration Price_per_Night Occupancy_Rate Customer_Rating Discount_Offered Total_Revenue;
    title "Correlation Matrix of Numerical Variables";
run;

/* Simple Reshape of Correlation Matrix for Heatmap */
data corr_matrix_long;
    set corr_matrix(where=(_TYPE_="CORR"));
    array vars[*] Stay_Duration Price_per_Night Occupancy_Rate Customer_Rating Discount_Offered Total_Revenue;
    do i = 1 to dim(vars);
        RowVar = _NAME_;
        ColVar = vname(vars[i]);
        CorrValue = vars[i];
        output;
    end;
    keep RowVar ColVar CorrValue;
```

```
run;

/* Plot the Heatmap */
proc sgplot data=corr_matrix_long;
    heatmapparm x=RowVar y=ColVar colorresponse=CorrValue / colormodel=(blue white red);
    gradlegend / title="Correlation";
    title "Correlation Heatmap of Numerical Variables";
run;
title;

/* ANOVA for Total_Revenue by Property_Type */
proc anova data=airbnb_imputed;
    class Property_Type;
    model Total_Revenue = Property_Type;
    title "ANOVA for Total Revenue by Property Type";
run;
title;

/* Create a binary variable based on Total_Revenue (high vs low) */
data airbnb_imputed;
    set airbnb_imputed;
    if Total_Revenue > 500 then Revenue_Category = 'High';
    else Revenue_Category = 'Low';
run;

/* Chi-Square Test for Property_Type by Revenue_Category */
proc freq data=airbnb_imputed;
    tables Property_Type * Revenue_Category / chisq;
    title "Chi-Square Test for Property Type by Revenue Category";
run;
title;

/* Simple Linear Regression - Predicting Total Revenue based on Price_per_Night and Occupancy_Rate */
proc reg data=airbnb_imputed;
    model Total_Revenue = Price_per_Night Occupancy_Rate;
    title "Linear Regression Model for Total Revenue";
run;
title;

/* Logistic Regression - Example */
proc logistic data=airbnb_imputed;
    model Occupancy_Rate(event='1') = Price_per_Night Customer_Rating Discount_Offered;
    title "Logistic Regression Model Example";
run;
***
```