```
from google.colab import files
uploaded = files.upload()
```

Choose Files Electric_Ve…n_Data.xlsx

- **Electric_Vehicle_Population_Data.xlsx**(application/vnd.openxmlformats-officedocument.spreadsheetml.sheet) - 21998974 bytes, last modified: 12/24/2024 - 100% done

Saving Electric_Vehicle_Population_Data.xlsx to Electric_Vehicle_Population_Data.xlsx

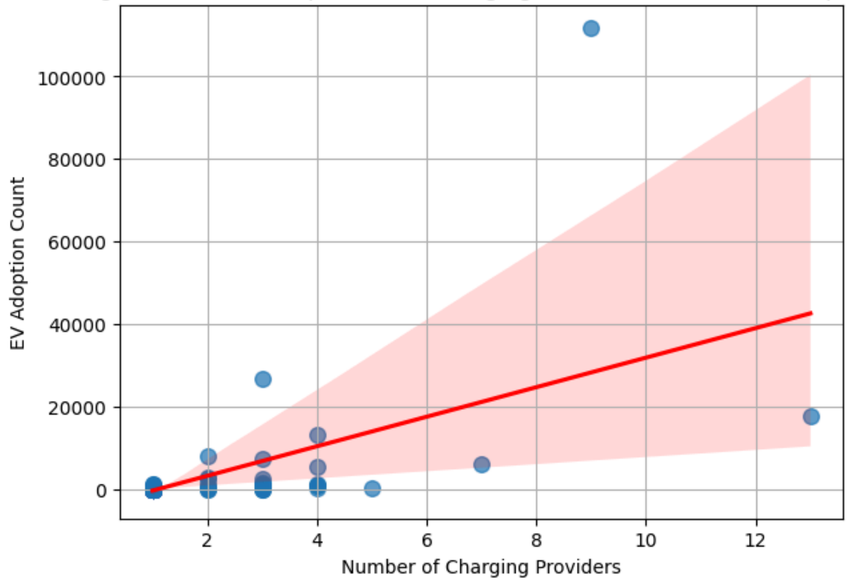
```
# Import Required Libraries
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import statsmodels.api as sm
from scipy.stats import spearmanr, kendalltau
from sklearn.cluster import KMeans

# Load Dataset (Already Uploaded by User)
df = pd.read_excel("Electric_Vehicle_Population_Data.xlsx")

# Extract necessary columns
columns_needed = ["County", "Make", "Model", "Electric Vehicle Type", "Electric Range", "Base MSRP", "Electric Utility"]
df = df[columns_needed]

# Aggregate data at the county level
county_data = df.groupby("County").agg(
    EV_Adoption=("Make", "count"),
    Charging_Providers=("Electric Utility", "nunique")
).dropna()

# Regression Analysis
# Prepare independent and dependent variables
X = sm.add_constant(county_data["Charging_Providers"])
y = county_data["EV_Adoption"]

# Fit linear regression model
model = sm.OLS(y, X).fit()

# Polynomial Regression
county_data["Charging_Providers^2"] = county_data["Charging_Providers"] ** 2
X_poly = sm.add_constant(county_data[["Charging_Providers", "Charging_Providers^2"]])
model_poly = sm.OLS(y, X_poly).fit()

# Logarithmic Regression
county_data["Log_Charging_Providers"] = np.log1p(county_data["Charging_Providers"])
X_log = sm.add_constant(county_data["Log_Charging_Providers"])
model_log = sm.OLS(y, X_log).fit()

# Correlation Analysis
spearman_corr, spearman_p = spearmanr(county_data["Charging_Providers"], county_data["EV_Adoption"])
kendall_corr, kendall_p = kendalltau(county_data["Charging_Providers"], county_data["EV_Adoption"])

# Clustering Analysis (K-Means)
kmeans = KMeans(n_clusters=3, random_state=42)
county_data["Cluster"] = kmeans.fit_predict(county_data[["Charging_Providers", "EV_Adoption"]])

# Print Regression Summaries and Correlation Results
print("\nLinear Regression Summary:\n", model.summary())
print("\nPolynomial Regression Summary:\n", model_poly.summary())
print("\nLogarithmic Regression Summary:\n", model_log.summary())
print(f"\nSpearman Correlation: {spearman_corr}, P-Value: {spearman_p}")
print(f"Kendall Correlation: {kendall_corr}, P-Value: {kendall_p}")
print("\nK-Means Cluster Centers:\n", kmeans.cluster_centers_)

# Visualizations
# Figure 1: Scatter Plot with Regression Line
plt.figure(figsize=(7, 5))
sns.regplot(x=county_data["Charging_Providers"], y=county_data["EV_Adoption"],
            scatter_kws={"s": 70, "alpha": 0.7}, line_kws={"color": "red"})
plt.xlabel("Number of Charging Providers")
plt.ylabel("EV Adoption Count")
plt.title("Figure 1: Relationship Between Charging Infrastructure and EV Adoption")
plt.grid(True)
plt.show()

# Figure 2: Bar Chart of Top 10 Counties by EV Adoption
```

```
# Figure 2: Bar Chart of Top 10 Counties by EV Adoption
plt.figure(figsize=(7, 5))
top_counties = county_data.sort_values("EV_Adoption", ascending=False).head(10)
sns.barplot(x=top_counties["EV_Adoption"], y=top_counties.index, palette="Blues_r")
plt.xlabel("Total EV Adoption")
plt.ylabel("County")
plt.title("Figure 2: Top 10 Counties by EV Adoption")
plt.grid(axis="x")
plt.show()

# Figure 3: Line Chart - EV Growth vs. Infrastructure Expansion
plt.figure(figsize=(7, 5))
sorted_df = county_data.sort_values("Charging_Providers")
plt.plot(sorted_df["Charging_Providers"], sorted_df["EV_Adoption"], marker="o", linestyle="-", color="blue")
plt.xlabel("Number of Charging Providers")
plt.ylabel("EV Adoption Count")
plt.title("Figure 3: EV Adoption Growth as Charging Infrastructure Expands")
plt.grid(True)
plt.show()

# Figure 4: Clustered Scatter Plot (K-Means Clustering)
plt.figure(figsize=(7, 5))
sns.scatterplot(x=county_data["Charging_Providers"], y=county_data["EV_Adoption"], hue=county_data["Cluster"],
                palette="coolwarm", s=80)
plt.xlabel("Number of Charging Providers")
plt.ylabel("EV Adoption Count")
plt.title("Figure 4: County Segmentation Based on EV Adoption & Charging Infrastructure")
plt.grid(True)
plt.show()
```

```
Linear Regression Summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:            EV_Adoption   R-squared:                       0.320
Model:                            OLS   Adj. R-squared:                  0.317
Method:                 Least Squares   F-statistic:                     97.13
Date:                Fri, 07 Feb 2025   Prob (F-statistic):           5.11e-19
Time:                        19:48:10   Log-Likelihood:                -2127.0
No. Observations:                 208   AIC:                             4258.
Df Residuals:                     206   BIC:                             4265.
Df Model:                           1
Covariance Type:            nonrobust
======================================================================================
                         coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------
const              -3931.4117    687.780     -5.716      0.000   -5287.402   -2575.422
Charging_Providers  3579.1574    363.171      9.855      0.000    2863.148    4295.167
==============================================================================
Omnibus:                      368.968   Durbin-Watson:                   1.997
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           118575.362
Skew:                           9.045   Prob(JB):                         0.00
Kurtosis:                     118.562   Cond. No.                         3.27
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Polynomial Regression Summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:            EV_Adoption   R-squared:                       0.323
Model:                            OLS   Adj. R-squared:                  0.316
Method:                 Least Squares   F-statistic:                     48.84
Date:                Fri, 07 Feb 2025   Prob (F-statistic):           4.51e-18
Time:                        19:48:10   Log-Likelihood:                -2126.6
No. Observations:                 208   AIC:                             4259.
Df Residuals:                     205   BIC:                             4269.
Df Model:                           2
Covariance Type:            nonrobust
==========================================================================================
                           coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------------
const                -3202.9488   1112.269     -2.880      0.004   -5395.902   -1009.996
Charging_Providers    2869.3102    925.727      3.100      0.002    1044.144    4694.476
Charging_Providers^2    72.8356     87.360      0.834      0.405     -99.404     245.075
==============================================================================
Omnibus:                      367.010   Durbin-Watson:                   1.999
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           118817.349
Skew:                           8.937   Prob(JB):                         0.00
Kurtosis:                     118.716   Cond. No.                         42.9
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Logarithmic Regression Summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:            EV_Adoption   R-squared:                       0.243
Model:                            OLS   Adj. R-squared:                  0.239
Method:                 Least Squares   F-statistic:                     66.08
Date:                Fri, 07 Feb 2025   Prob (F-statistic):           3.98e-14
Time:                        19:48:10   Log-Likelihood:                -2138.2
No. Observations:                 208   AIC:                             4280.
Df Residuals:                     206   BIC:                             4287.
Df Model:                           1
Covariance Type:            nonrobust
==========================================================================================
                             coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------------
const                  -9506.2722   1389.437     -6.842      0.000   -1.22e+04   -6766.932
Log_Charging_Providers  1.308e+04   1609.436      8.129      0.000    9909.987    1.63e+04
==============================================================================
Omnibus:                      392.709   Durbin-Watson:                   1.988
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           153222.971
Skew:                          10.231   Prob(JB):                         0.00
Kurtosis:                     134.381   Cond. No.                         5.54
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Spearman Correlation: 0.6143675296448879, P-Value: 5.659801759263228e-23
```

```
Kendall Correlation: 0.5347362923935711, P-value: 3.3060004746161486e-19

K-Means Cluster Centers:
 [[1.27941176e+00 2.50127451e+02]
 [9.00000000e+00 1.11709000e+05]
 [6.66666667e+00 1.91623333e+04]]
```

Figure 1: Relationship Between Charging Infrastructure and EV Adoption




```
<ipython-input-2-0591f0bcc46d>:71: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` variable to `hue` and set `lege

  sns.barplot(x=top_counties["EV_Adoption"], y=top_counties.index, palette="Blues_r")
```

Figure 2: Top 10 Counties by EV Adoption

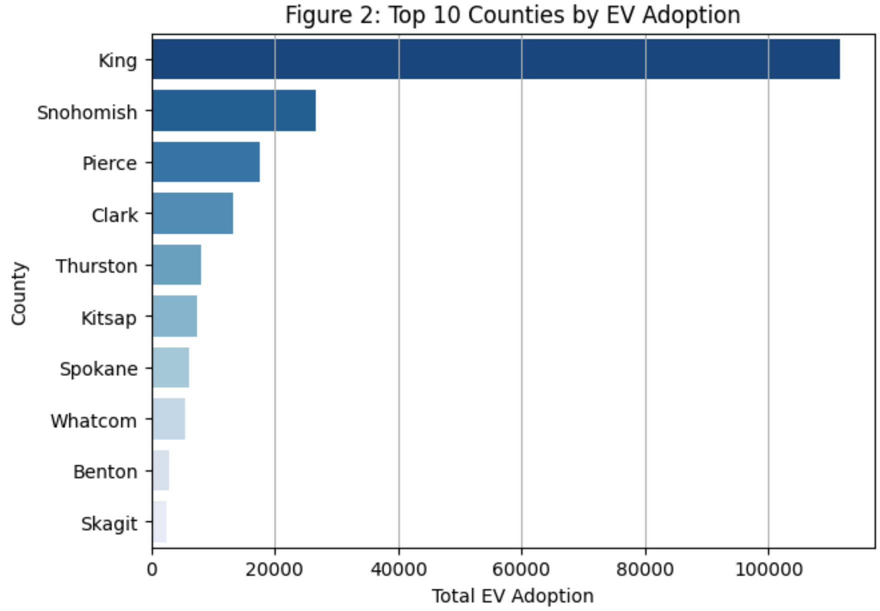
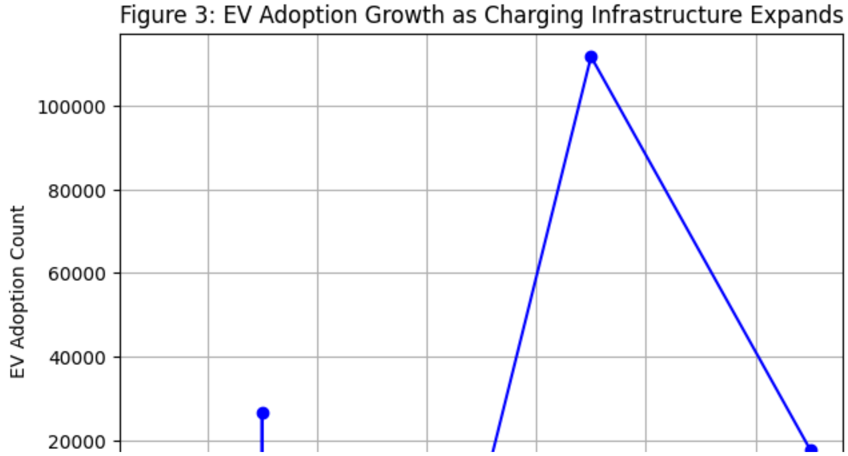
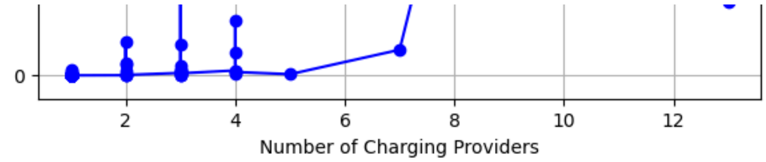


Figure 3: EV Adoption Growth as Charging Infrastructure Expands




0
2
4
6
8
10
12
Number of Charging Providers