STRATEGY FOR EV AND EV CHARGING STATIONS

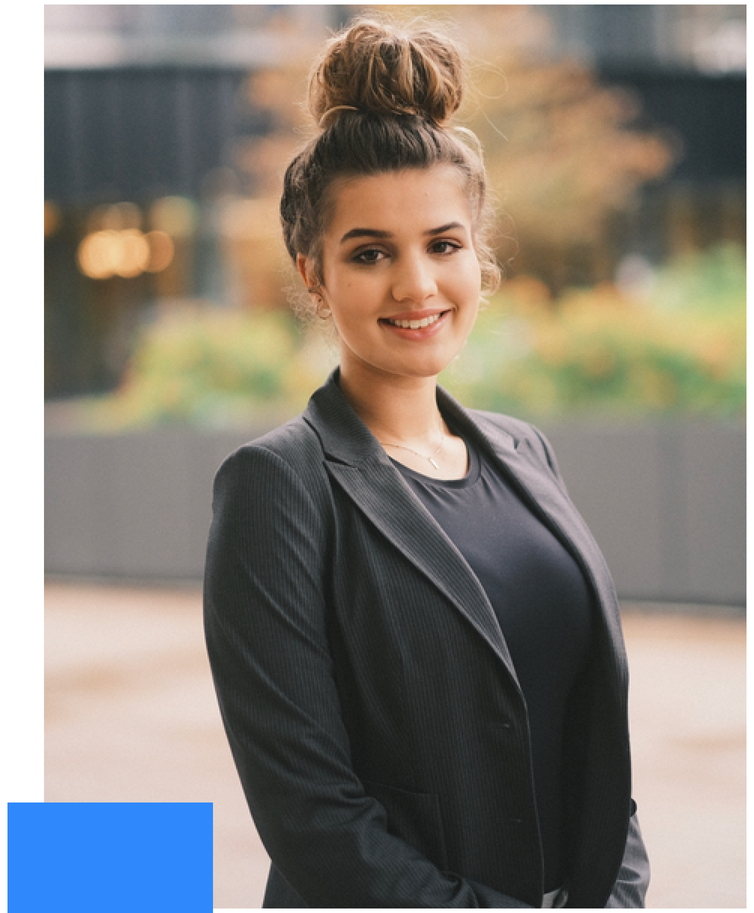
General Motors

Presented by
Jack Hu,
Haoxiang Yi,
Jaida Adamson,
Boyuan Lu

OUR TEAM

JACK HU

HAOXIANG YI



JAIDA ADAMSON

BOYUAN LU

TABLE OF CONTENT

BACKGROUND i

APPROACH ii

DATA COLLECTION AND TRANSFORMATION iii

EV ADOPTION RATE MODELING iv

CHARGING STATIONS MODELING v

MARGINAL ANALYSIS vi

CONCLUSION vii

BACKGROUND

EXPLORATION OF VARIABLES INFLUENCING ELECTRIC VEHICLE (EV) ADOPTION AND CHARGER PLACEMENT.

Key Questions to Answer:

1. What are the factors that contribute to EV adoption, and what role does charging infrastructure play?
2. What areas (zip codes) should see the most growth if new EV chargers are installed?
3. Where specifically (venues) within the highest priority areas should EV chargers be installed?

APPROACH

Defining the target variables:

EV Adoption Rate:

- Predict based on all useful variables
- Identify most significant variables affecting adoption
- Determine the best ways to promote EVs

APPROACH

Defining the target variables:

Charging Stations Count:

- Predict based on all useful variables
- Identify the most significant variables affecting number of charging stations and use it determine the most influential venues
- Focus on the top 100 zip codes with the largest differences between predicted and actual
- Analyze the top 100 zip codes and find the best way to allocate the 100 charging stations based on the marginal increase in EV adoption rate

DATA COLLECTION

DATA SOURCES:

1. WEATHER INFORMATION
   - AVERAGE PRECIPITATION, TEMPERATURE PER ZIP CODE

3. TOURIST-RELATED INFORMATION
   - NUMBER OF TOURISTS, HOTELS PER ZIP CODE/COUNTY

4. TRAFFIC-RELATED DATA
   - NUMBER OF MAJOR ROADS, TRANSIT PLACES PER ZIP CODE/COUNTY

5. CHARGING STATIONS
   - LOCATIONS AND COUNTS OF PUBLIC AND PRIVATE CHARGING STATIONS

6. VENUE DATA
   -NUMBER OF RETAIL, OFFICES, PARKING, MEDICAL/EDUCATIONAL PER ZIP CODE/COUNTY

DATA TRANSFORMATION

01 Data Cleaning and Preprocessing
- Grouped data by county or zip code level
- Combined main data provided by the client with additional influential data

02 Missing Value Handling
- Grouped by county and found the average to fill in some missing variables
- Created dummy variables for categorical variables

03 Data Integration
- Filtered data for the chosen state: **Texas**
- Ensured consistency when combining datasets

04 Feature Engineering
- Developed new columns to better explain target variables, such as "charging stations per EV"

EV ADOPTION RATE MODELING

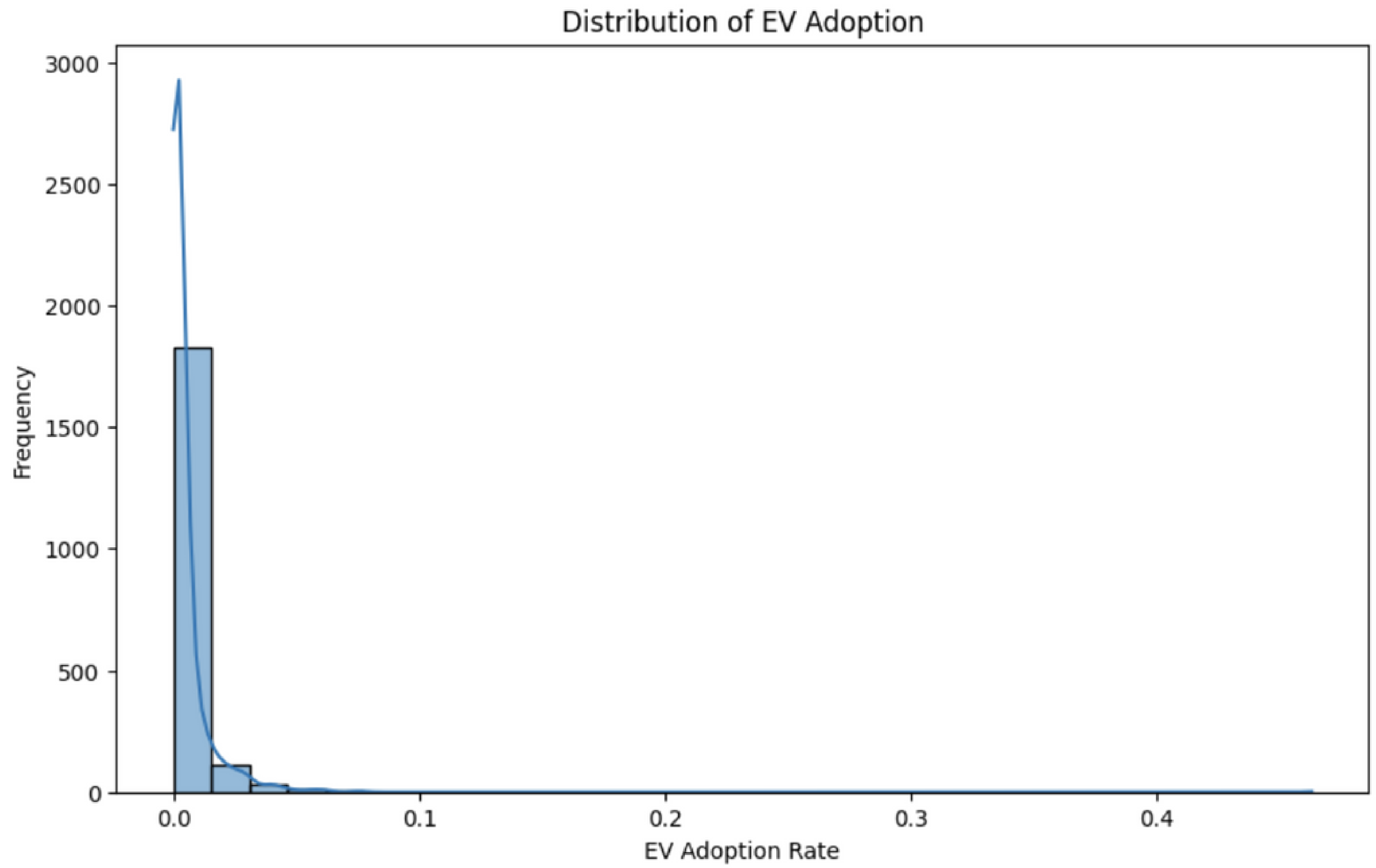
EDA

LINEAR REGRESSION

RANDOM FOREST

MODEL EVALUATION

EDA

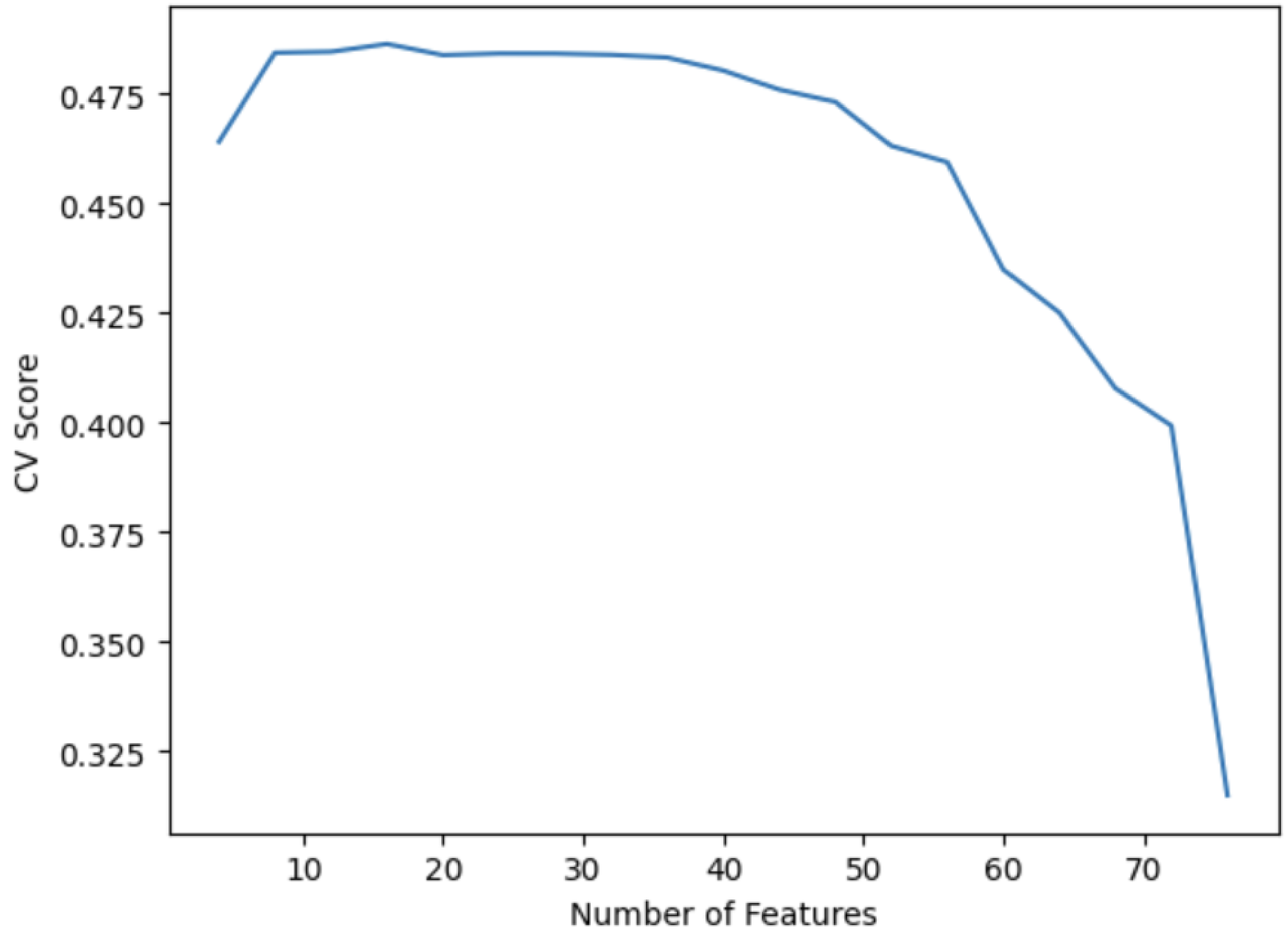


Highly skewed distribution

Adoption rate of EVs is generally low

Applying a logarithmic scale to the data

LINEAR REGRESSION



The CV score starts at a relatively high value when the number of features is small (near 4), which suggests that a few features are already providing a significant amount of predictive power.

The most significant drop happens after approximately 50 features

Optimal number of features for linear regression: 16

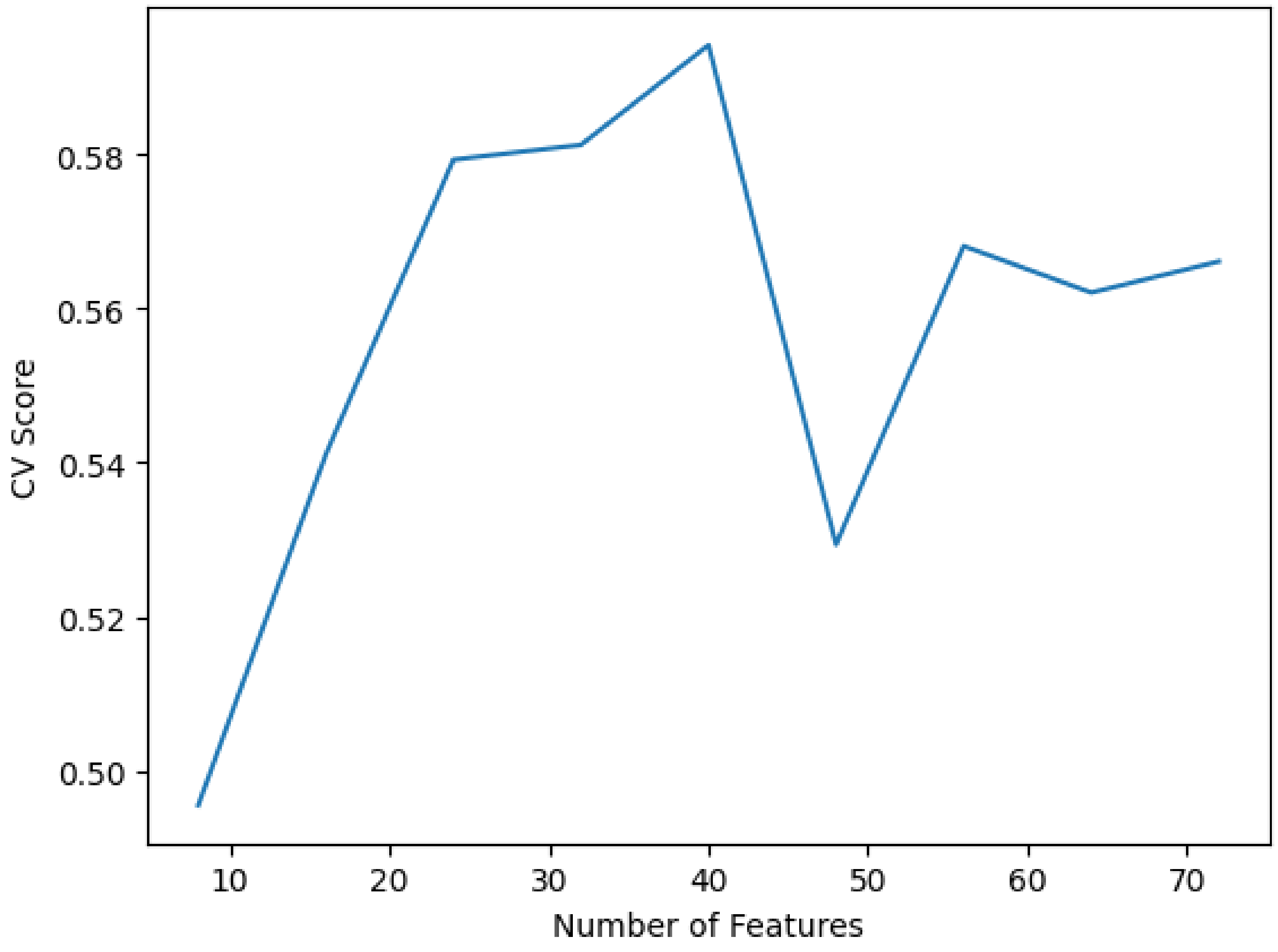
LINEAR REGRESSION

wfh_rate	2.670000e-02
Total_EV_Chargers	2.000000e-04
Total_Hotel_Airbnb	8.789000e-07
individual_income_percapita	1.192000e-07
household_income_median	5.988000e-08
RUCC_2013	-3.000000e-04

R-squared value: 0.7321320282984156

Mean squared error (MSE) is 1.59e-05

RANDOM FOREST

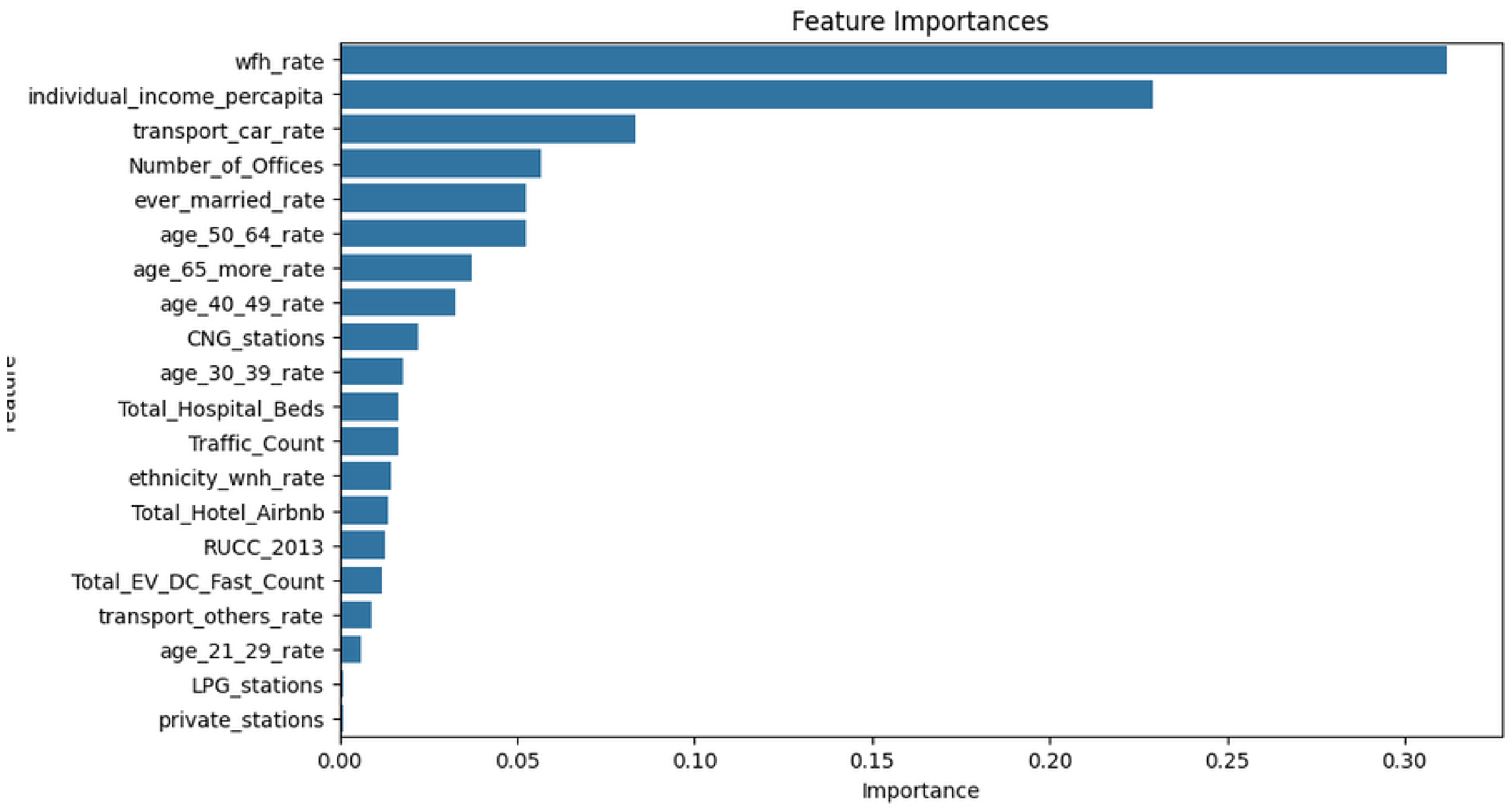


Optimal number of features for random forest: 40

cv score of 0.5942108029867018

R-squared value for training data: 0.8884544338686554

MODEL EVALUATION



	Linear Regression	Random Forest
R-squared	0.7321	0.73207
MSE	1.590716e-05	1.591030e-05

- Performance: Both models perform equally well on the test set.
- Complexity: The Linear Regression model is generally simpler and faster to train than a Random Forest, and it offers the advantage of interpretability.
- Deployment: For deployment considerations, simplicity and speed can be advantageous.

MODEL COMPARISON

Overlapping features
RUCC_2013
transport_car_rate
wfh_rate
individual_income_percapita
LPG_stations
Total_Hotel_Airbnb

- **Policy Implications:**
  - RUCC_2013
  - Transport_car_rate
  - WFH_rate
- **Economic Considerations**:
  - Individual_income_percapita
  - LPG_stations
- **Tourism**
  - Total_Hotel_Airbnb

- **These overlapping features can help in crafting nuanced approaches to promote EVs, considering the diversity in demographics and economic conditions.**

NUMBER OF CHARGERS MODELING

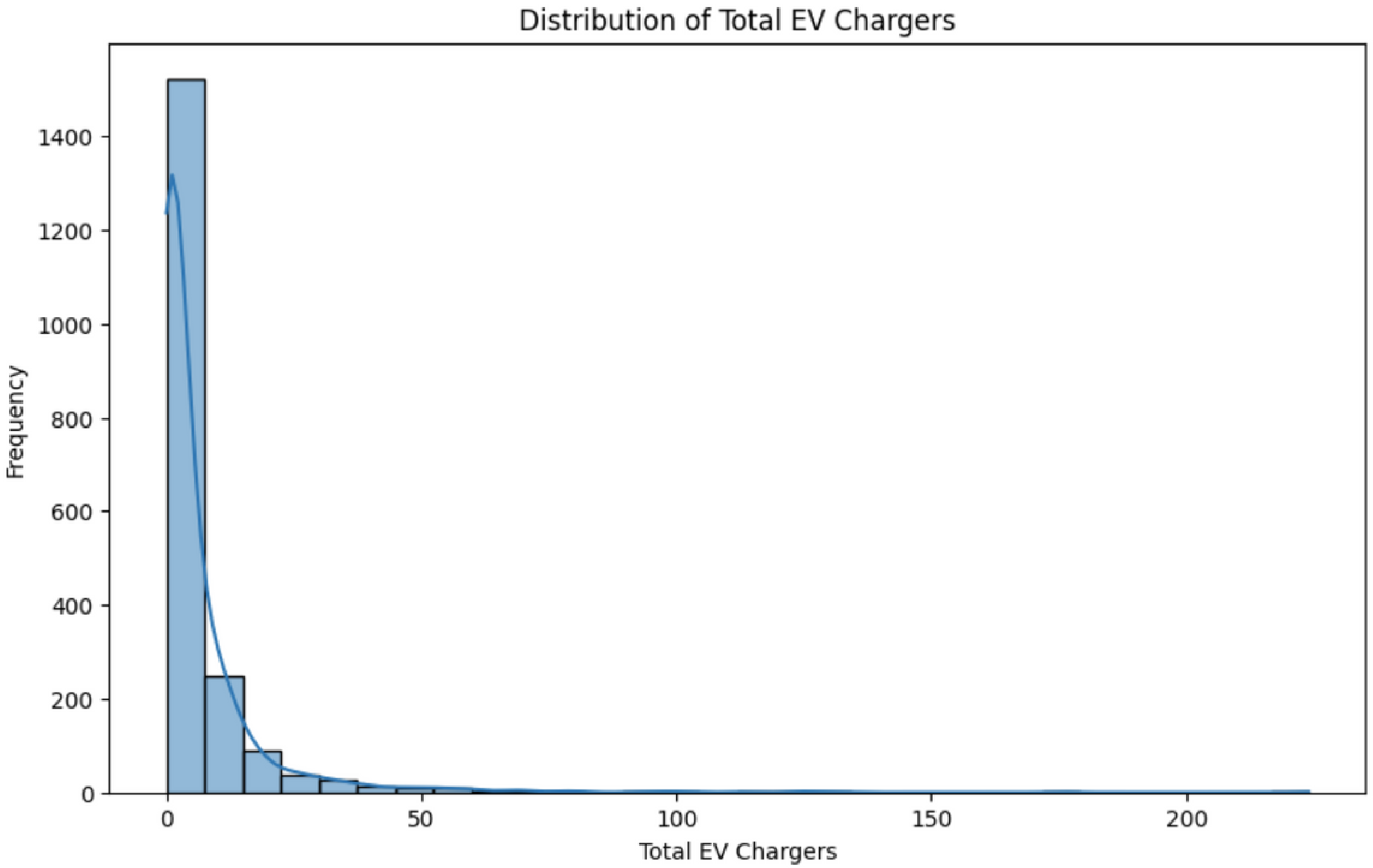
EDA

LINEAR REGRESSION

RANDOM FOREST

MODEL EVALUATION

EDA

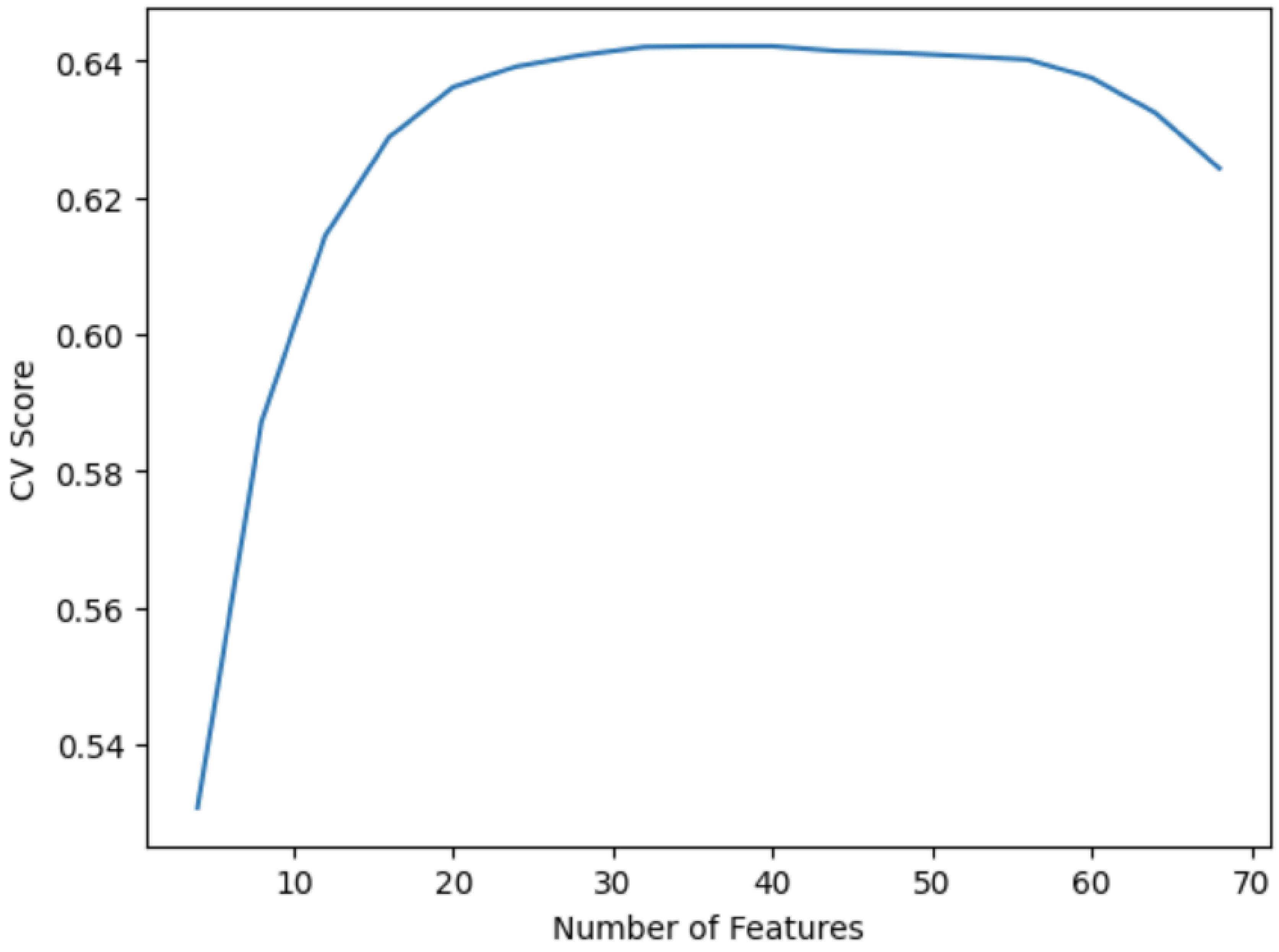


Highly skewed distribution

The number of EV Chargers is low

Applying logarithmic transformation

LINEAR REGRESSION



Optimal number of features for linear regression: 36

LINEAR REGRESSION

	coef	std err	t	P>|t|	[0.025	0.975]
transport_public_rate	1.181300e+01	2.656000e+00	4.447	0.000	6.603000e+00	1.702300e+01
Most_Common_Charging_Facility_Type_TRAVEL_CENTER	1.949900e+00	3.290000e-01	5.921	0.000	1.304000e+00	2.596000e+00
age_21_29_rate	1.868000e+00	2.680000e-01	6.974	0.000	1.343000e+00	2.393000e+00
Most_Common_Charging_Facility_Type_PAY_GARAGE	1.836900e+00	6.580000e-01	2.790	0.005	5.460000e-01	3.128000e+00
Most_Common_Charging_Facility_Type_PUBLIC	1.827400e+00	4.650000e-01	3.928	0.000	9.150000e-01	2.740000e+00
Most_Common_Charging_Facility_Type_AIRPORT	1.806100e+00	5.170000e-01	3.494	0.000	7.920000e-01	2.820000e+00
Most_Common_Charging_Facility_Type_GROCERY	1.679800e+00	4.660000e-01	3.603	0.000	7.650000e-01	2.594000e+00
Most_Common_Charging_Facility_Type_GAS_STATION	1.500900e+00	1.390000e-01	10.777	0.000	1.228000e+00	1.774000e+00
age_30_39_rate	1.313500e+00	3.360000e-01	3.913	0.000	6.550000e-01	1.972000e+00
Most_Common_Charging_Facility_Type_OTHER	8.589000e-01	1.510000e-01	5.704	0.000	5.640000e-01	1.154000e+00
Most_Common_Charging_Facility_Type_HOSPITAL	8.161000e-01	3.820000e-01	2.136	0.033	6.700000e-02	1.566000e+00
Most_Common_Charging_Facility_Type_CAR_DEALER	7.933000e-01	9.700000e-02	8.200	0.000	6.040000e-01	9.830000e-01
Most_Common_Charging_Facility_Type_SHOPPING_CENTER	7.725000e-01	1.530000e-01	5.036	0.000	4.720000e-01	1.073000e+00
Most_Common_Charging_Facility_Type_FED_GOV	7.633000e-01	2.100000e-01	3.630	0.000	3.510000e-01	1.176000e+00

28 significant variables

Testing R-squared value of 0.718

Mean Squared Error (MSE) is 0.621

LINEAR REGRESSION

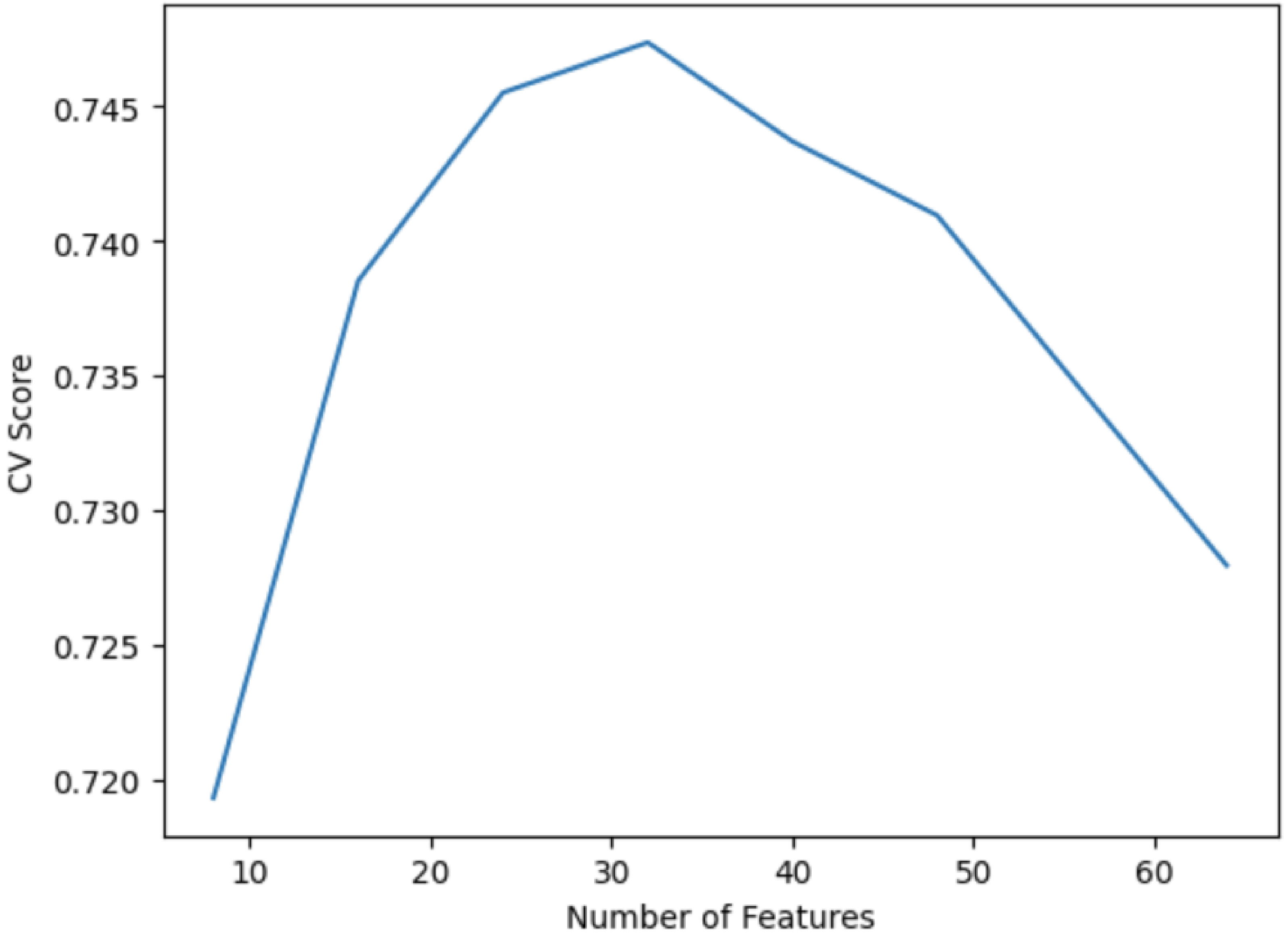
	coef	std err	t	P>|t|	[0.025	0.975]
age_0_20_rate	7.304000e-01	1.830000e-01	3.982	0.000	3.710000e-01	1.090000e+00
Most_Common_Charging_Facility_Type_HOTEL	7.272000e-01	6.500000e-02	11.251	0.000	6.000000e-01	8.540000e-01
Most_Common_Charging_Facility_Type_MUNI_GOV	6.565000e-01	3.300000e-01	1.990	0.047	9.000000e-03	1.304000e+00
Most_Common_Charging_Facility_Type_OFFICE_BLDG	5.472000e-01	2.110000e-01	2.594	0.010	1.330000e-01	9.610000e-01
age_65_more_rate	4.669000e-01	2.120000e-01	2.197	0.028	5.000000e-02	8.840000e-01
Most_Common_Charging_Facility_Type_PARKING_LOT	4.398000e-01	2.200000e-01	1.995	0.046	7.000000e-03	8.720000e-01
three_year_precipitation_avg	4.810000e-02	1.800000e-02	2.659	0.008	1.300000e-02	8.400000e-02
Total_EV	4.000000e-04	7.720000e-05	5.307	0.000	0.000000e+00	1.000000e-03
Total_Hotel_Airbnb	2.000000e-04	3.820000e-05	5.402	0.000	0.000000e+00	0.000000e+00
individual_income_percapita	1.135000e-05	1.470000e-06	7.708	0.000	8.460000e-06	1.420000e-05
total_population	6.179000e-06	1.420000e-06	4.342	0.000	3.390000e-06	8.970000e-06
Traffic_Count	-5.931000e-09	2.240000e-09	-2.652	0.008	-1.030000e-08	-1.540000e-09
RUCC_2013	-8.790000e-02	9.000000e-03	-10.145	0.000	-1.050000e-01	-7.100000e-02
ethnicity_wnh_rate	-3.099000e-01	9.200000e-02	-3.362	0.001	-4.910000e-01	-1.290000e-01

28 significant variables

Testing R-squared value of 0.718

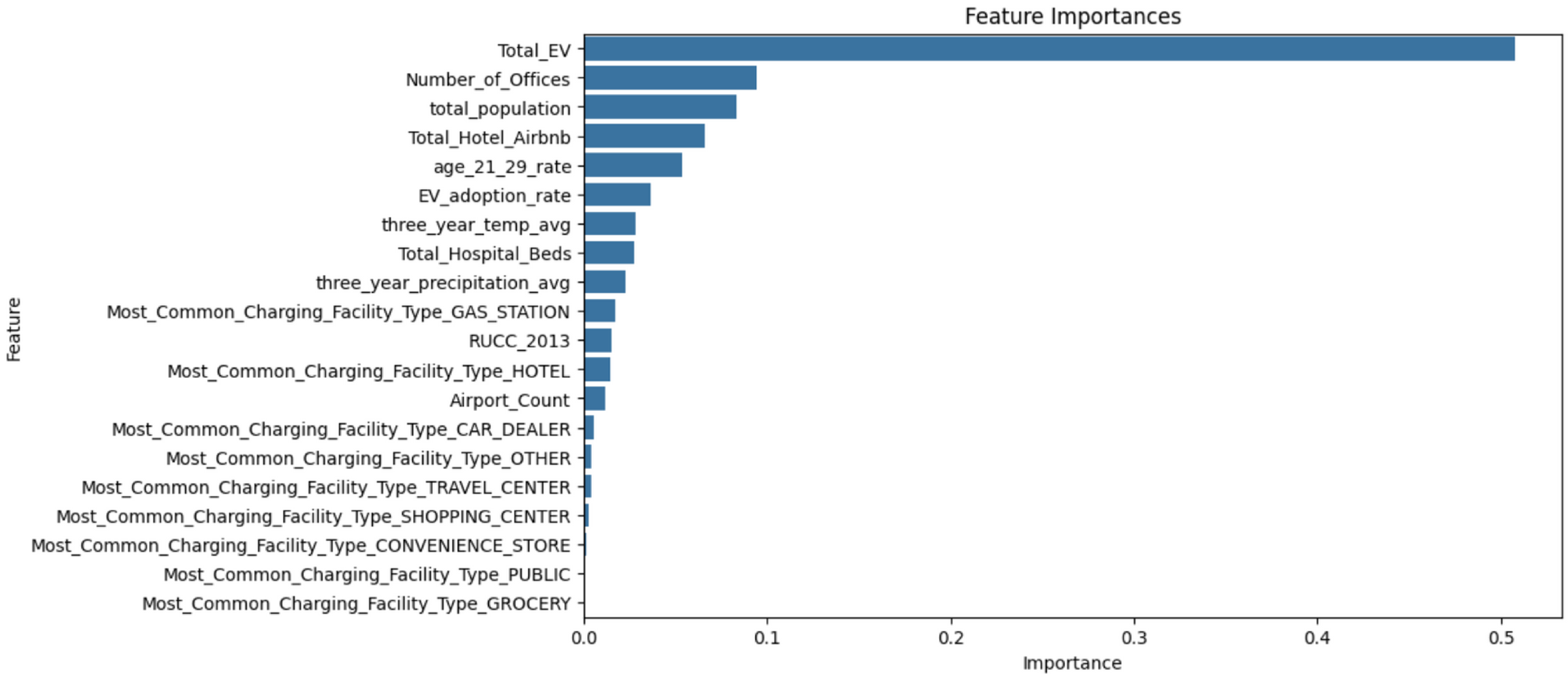
Mean Squared Error (MSE) is 0.621

RANDOM FOREST



Optimal number of features for random forest: 32

RANDOM FOREST



MODEL COMPARISON

RUCC_2013	Most_Common_Charging_Facility_Type_CAR_DEALER
total_population	Most_Common_Charging_Facility_Type_GAS_STATION
age_21_29_rate	Most_Common_Charging_Facility_Type_GROCERY
Total_EV	Most_Common_Charging_Facility_Type_HOTEL
Total_Hotel_Airbnb	Most_Common_Charging_Facility_Type_PUBLIC
three_year_precipitation_avg	Most_Common_Charging_Facility_Type_TRAVEL_CENTER
three_year_temp_avg	Most_Common_Charging_Facility_Type_SHOPPING_CENTER

MODEL COMPARISON

	R-Squared	MSE
Linear Regression	0.718	0.621
Random Forest	0.805	0.267

MODEL EVALUATION

Zip Code	Difference
78206	8
78731	7
78751	7
78256	6
77006	6
77003	5
75025	5
75251	5
77407	4
75039	4

- Using the Random Forest model, we predicted the number of EV charging stations that each zip code should have
- Difference = Prediction - Actual
- Top 10 zip codes with the largest difference is shown here

MARGINAL ANALYSIS

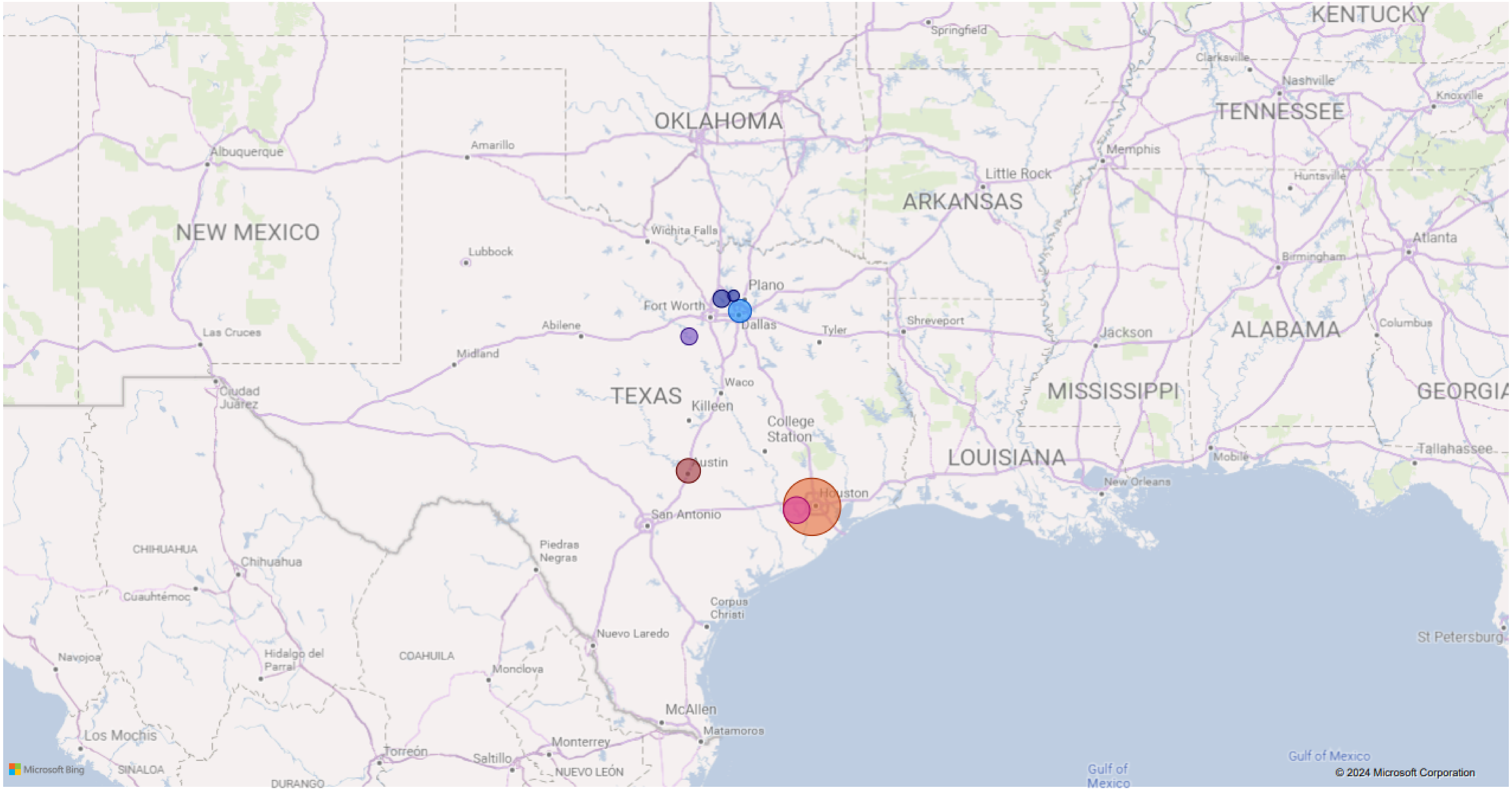
WHERE TO PUT THE CHARGERS?

We ran random forest models iteratively to determine which ZIP codes show the greatest increase in EV adoption rates when additional charging stations are allocated one at a time.

ZIP	County	Current_Number_of_Charging_Stations	Additional_Charging_Stations	Increase_in_EV_Adoption_Rate
77046	Harris	10	67	0.388479862
78751	Travis	9	10	0.049241123
75206	Dallas	5	9	0.043053624
77407	Fort Bend	10	7	0.063810303
75022	Denton	1	3	0.019884357
76049	Hood	3	3	0.017847721
75056	Denton	21	1	0.00455626

Increase_in_EV_Adoption_Rate by ZIP and County

County Dallas Denton Fort Bend Harris Hood Travis



WHERE TO PUT THE CHARGERS?

Do these locations make sense?

ZIP	County	Number_of_Charging_Stations_per_EV	RUCC_2013
77046	Harris	0.416666667	1
78751	Travis	0.027108434	1
75206	Dallas	0.006849315	1
77407	Fort Bend	0.006666667	1
75022	Denton	0.001081081	1
76049	Hood	0.015789474	1
75056	Denton	0.014037433	1

WHERE TO PUT THE CHARGERS?

What are the precise impacts?

9193 New EVs

$53,376* X 9193 =

$491M REVENUE

ZIP	County	Additional_Charging_Stations	Increase_in_EV_Adoption_Rate	Number_of_EVs_Added
77046	Harris	67	0.388479862	487.1537472
78751	Travis	10	0.049241123	803.9105747
75206	Dallas	9	0.043053624	1654.636864
77407	Fort Bend	7	0.063810303	4887.869181
75022	Denton	3	0.019884357	536.8776412
76049	Hood	3	0.017847721	522.9560853
75056	Denton	1	0.00455626	299.7335674

*According to the 2023 Kelly Blue Book

CONCLUSION

Q1 Factors Influencing EV Adoption Rate

- Urbanization
- Car Usage
- Life Style
- Income
- LPG Rate
- Hotel

Q2 Where To Put the Chargers?

- 77046
- 78751
- 75206
- 77407
- 75022
- 76049
- 75056
- $491M Revenue

Q3 What Venue Specifically to Put the Chargers?

- Car Dealership
- Gas Station
- Grocery
- Hotel
- Public Parking Lot
- Travel Center
- Shopping Center