PEAK PERFORMANCE:
PREDICTIVE MODELING FOR VACATION RENTAL REVENUE

CONTENT

01 BUSINESS PARTNER

02 BUSINESS OBJECTIVES

03 DATA ANALYSIS AND METHODS

04 MODELING

05 CONCLUSION

06 NEXT STEPS

07 CONTACTS

BUSINESS PARTNER

CLOUD 9 CABINS

SMOKY MOUNTAINS

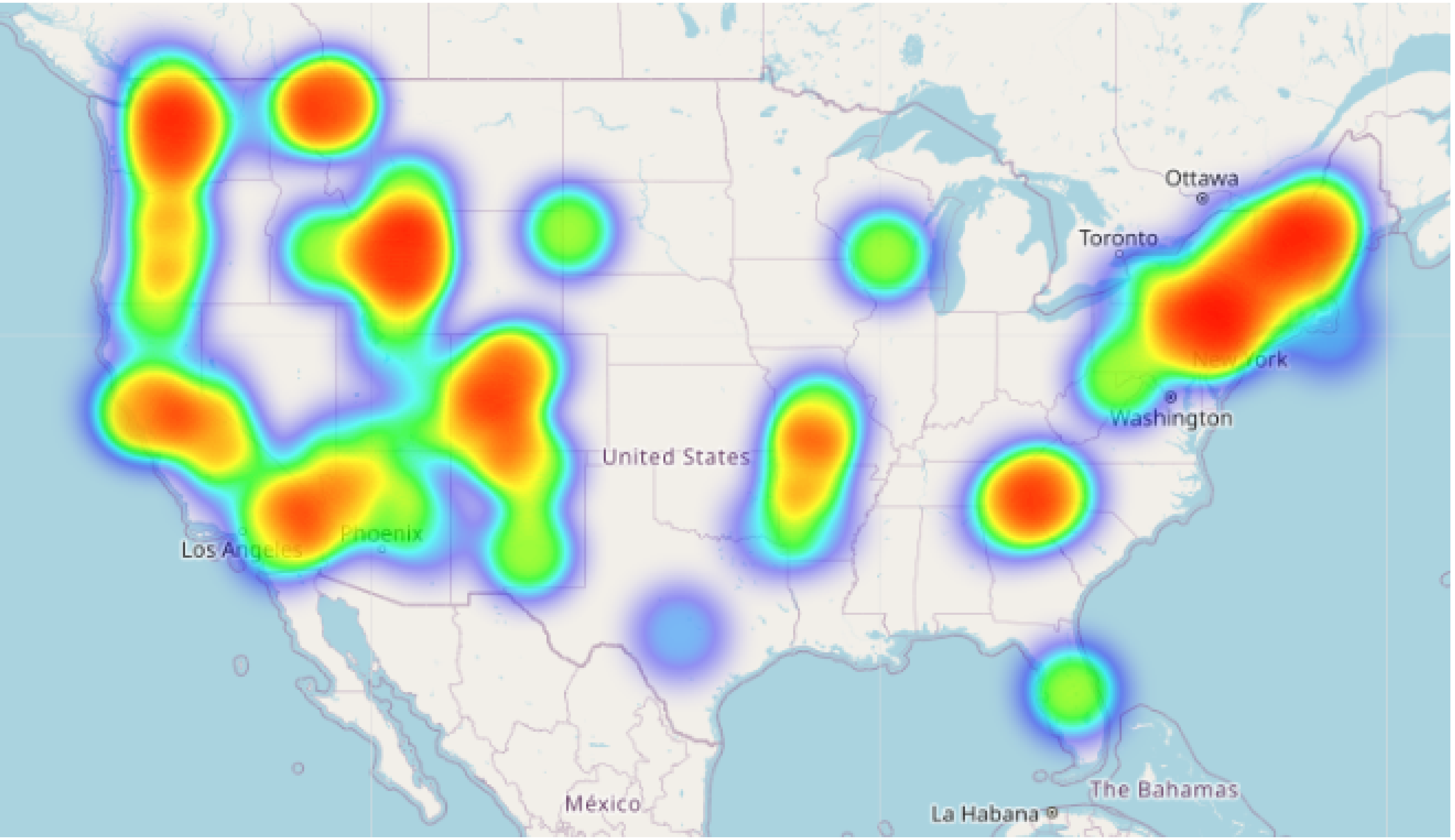
BUSINESS PROBLEM/OBJECTIVE

The business problem we aim to address is determining the profitability of potential new properties before committing significant resources to their development.

DATA ANALYSIS

AIRDNA

The Data

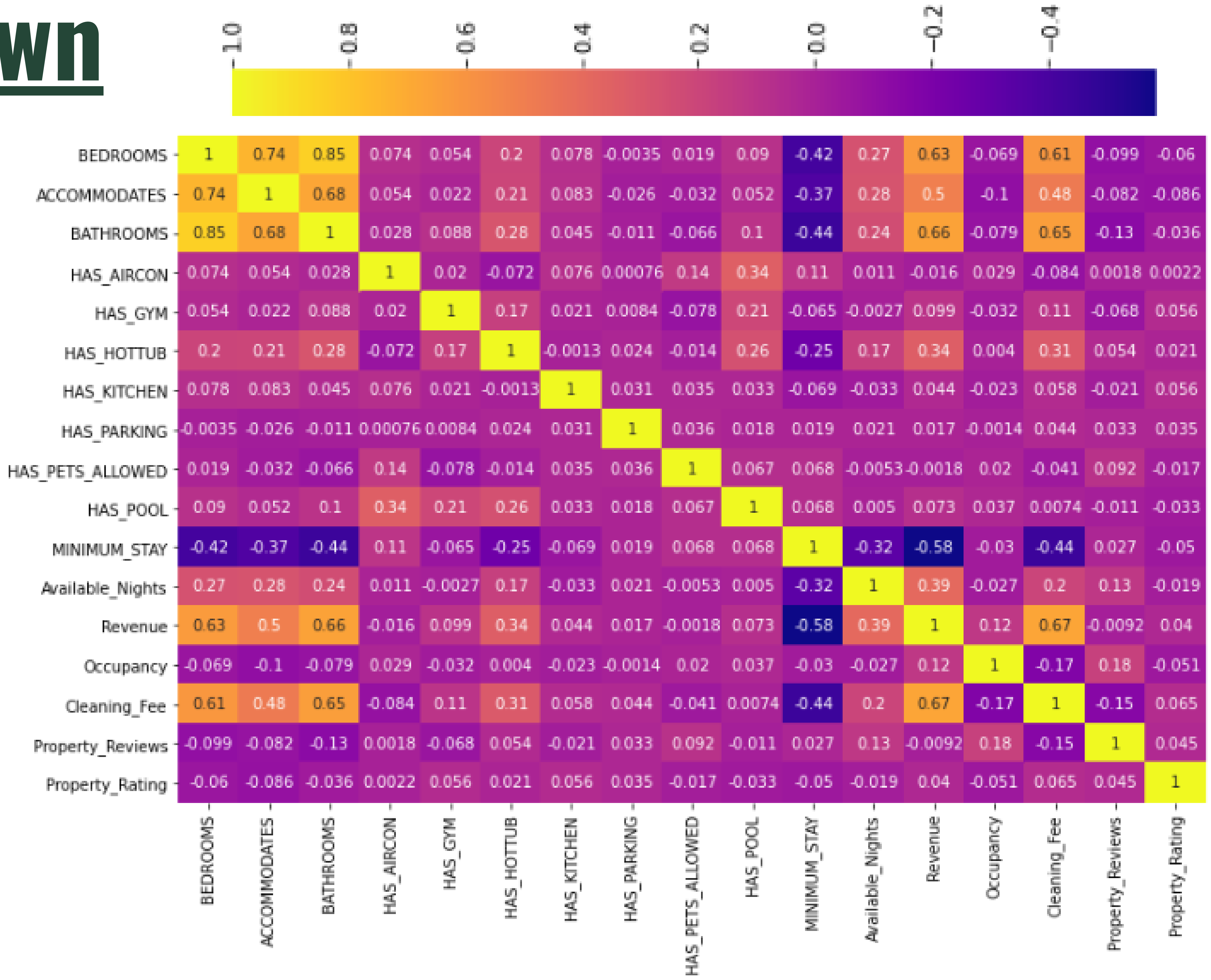



```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9039 entries, 0 to 9038
Data columns (total 33 columns):
 #   Column                     Non-Null Count  Dtype  
---  ------                     --------------  -----  
 0   Property ID                9039 non-null   object 
 1   TITLE                      9039 non-null   object 
 2   Property Manager/ Host ID  9039 non-null   object 
 3   BEDROOMS                   9029 non-null   float64
 4   ACCOMMODATES               9031 non-null   float64
 5   Airbnb Host URL            6397 non-null   object 
 6   Airbnb Listing URL         6397 non-null   object 
 7   BATHROOMS                  9035 non-null   float64
 8   CITY_NAME                  9038 non-null   object 
 9   HAS_AIRCON                 9039 non-null   bool   
 10  HAS_GYM                    9039 non-null   bool   
 11  HAS_HOTTUB                 9039 non-null   bool   
 12  HAS_KITCHEN                9039 non-null   bool   
 13  HAS_PARKING                9039 non-null   bool   
 14  HAS_PETS_ALLOWED           9039 non-null   bool   
 15  HAS_POOL                   9039 non-null   bool   
 16  INSTANT_BOOK               9022 non-null   object 
 17  LATITUDE                   9039 non-null   float64
 18  LONGITUDE                  9039 non-null   float64
 19  PRICE_TIER                 9039 non-null   object 
 20  STATE_NAME                 9039 non-null   object 
 21  SUPERHOST                  6064 non-null   object 
 22  Vrbo Listing URL           6104 non-null   object 
 23  ZIPCODE                    9039 non-null   int64  
 24  MINIMUM_STAY               8539 non-null   float64
 25  Available Nights           9039 non-null   int64  
 26  Revenue                    9039 non-null   int64  
 27  Revenue Potential          9039 non-null   float64
 28  ADR                        9018 non-null   float64
 29  Occupancy                  9039 non-null   float64
 30  Cleaning Fee               8269 non-null   float64
 31  Property Reviews           8334 non-null   float64
 32  Property Rating            7104 non-null   float64
dtypes: bool(7), float64(12), int64(3), object(11)
memory usage: 1.9+ MB
```

Narrowing Down the Data

Property Filtering Criteria

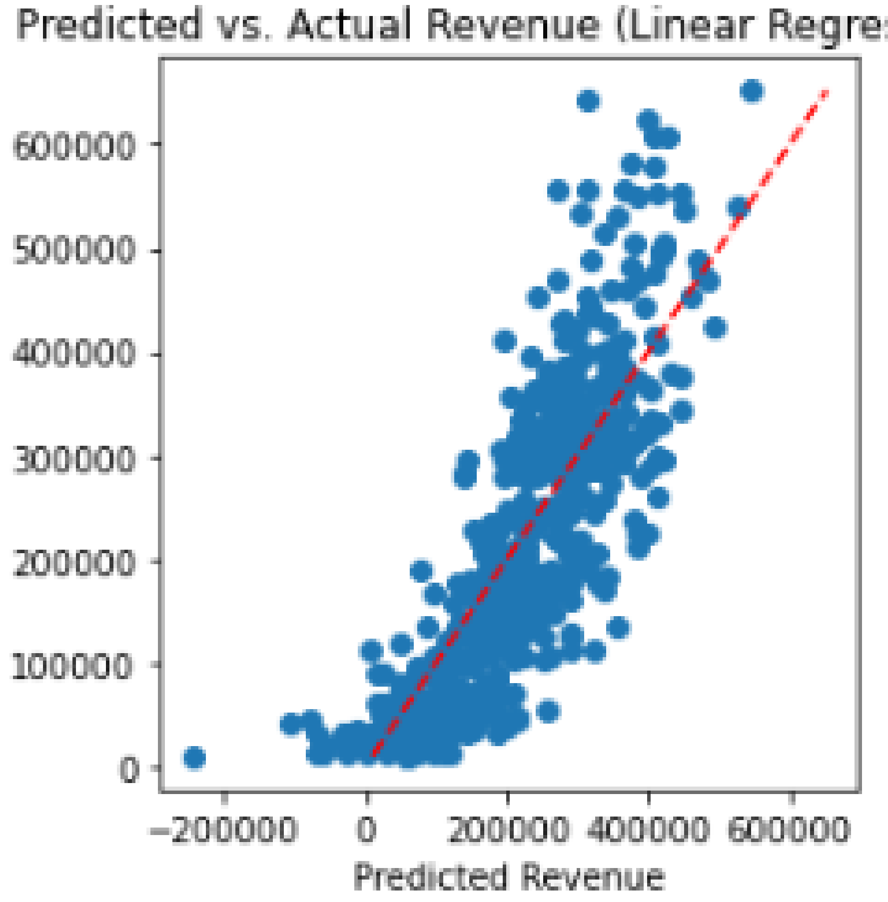
- Available Nights
  - Fewer than 200 nights: Likely not true short-term rentals.
- *Minimum Stay*
  - Over 100 days: Indicates long-term rentals.
- Property Reviews
  - Fewer than 10 reviews: Possibly unsuccessful or lack full data.
- Revenue
  - Less than $10,000: Potentially unsuccessful or temporary listings.
  - Over $700,000: Outliers.
- Occupancy
  - Below 25%: May indicate unsuccessful rentals or cost subsidization.

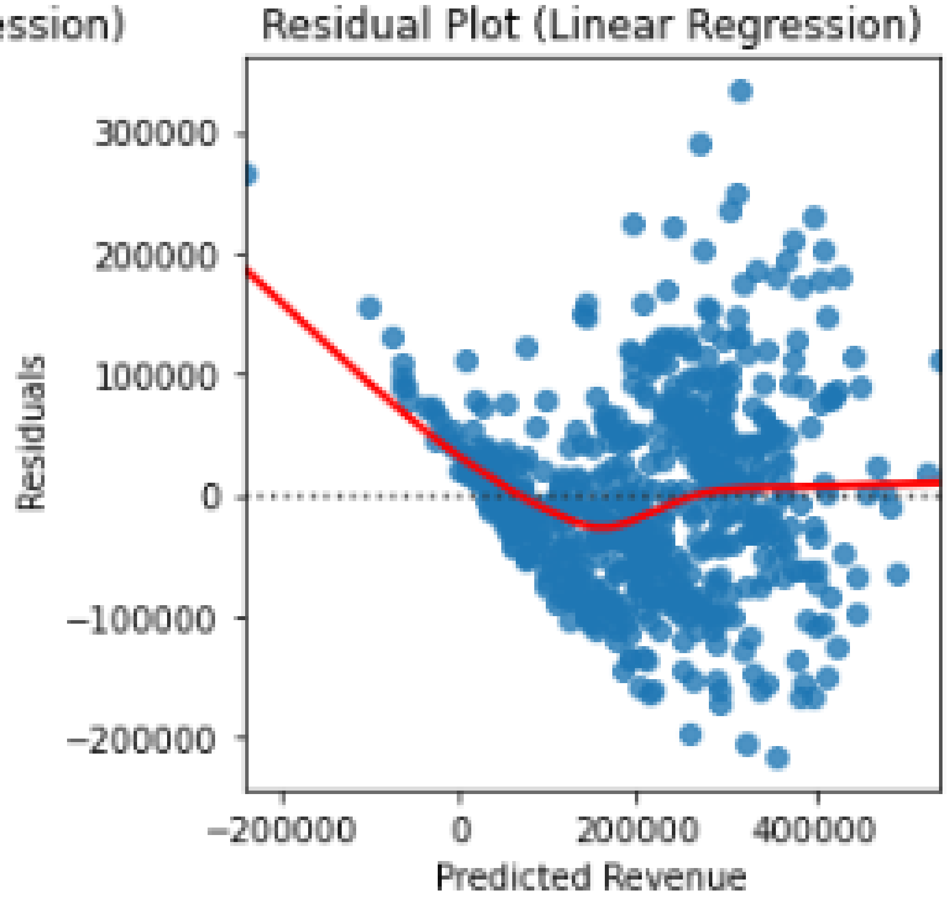


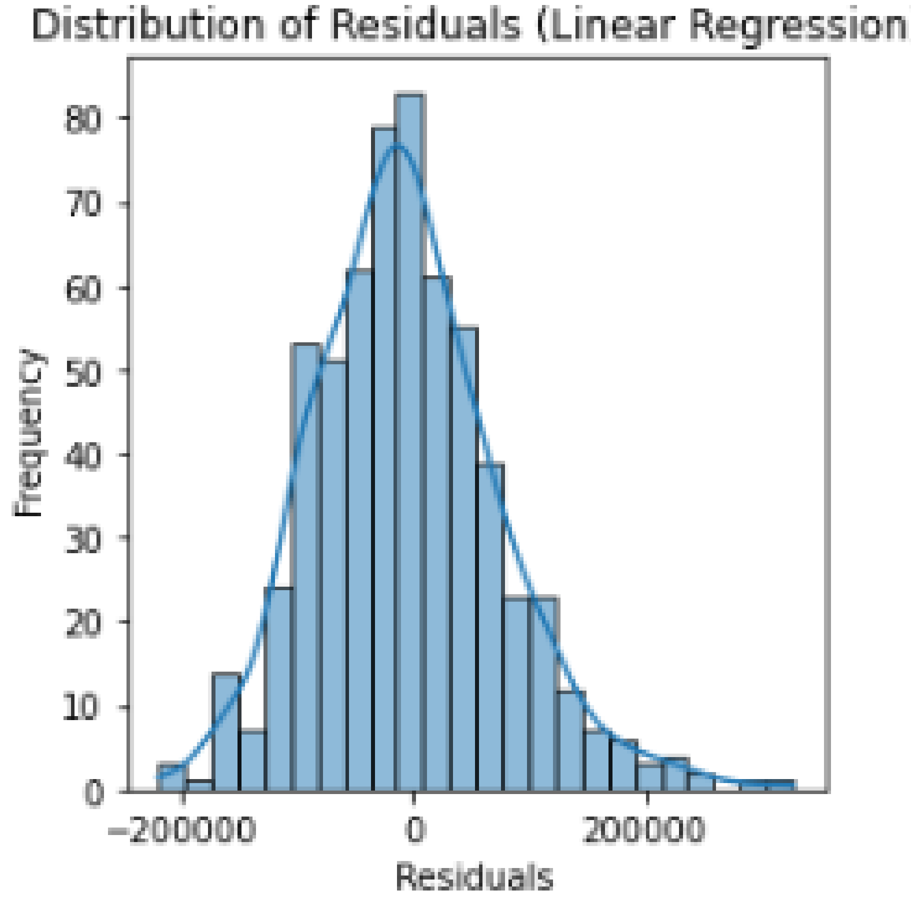
Final Data Set

```
ass 'pandas.core.frame.DataFrame'>
64Index: 3070 entries, 0 to 9035
a columns (total 13 columns):
  Column            Non-Null Count  Dtype
  ------            --------------  -----
  BEDROOMS          3070 non-null   float64
  ACCOMMODATES      3070 non-null   float64
  HAS_GYM           3070 non-null   int64
  HAS_HOTTUB        3070 non-null   int64
  HAS_KITCHEN       3070 non-null   int64
  HAS_PARKING       3070 non-null   int64
  HAS_PETS_ALLOWED  3070 non-null   int64
  HAS_POOL          3070 non-null   int64
  PRICE_TIER        3070 non-null   int64
  STATE_NAME        3070 non-null   object
  MINIMUM_STAY      3070 non-null   float64
  Available_Nights  3070 non-null   int64
  Cleaning_Fee      2902 non-null   float64
```

Baseline Model

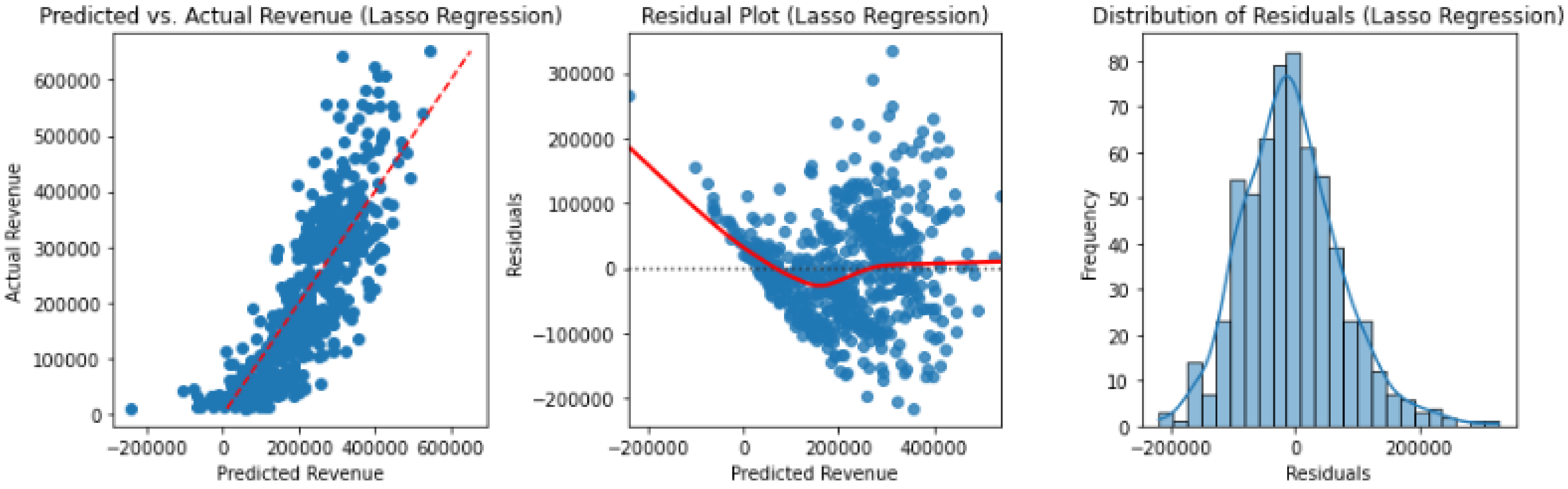




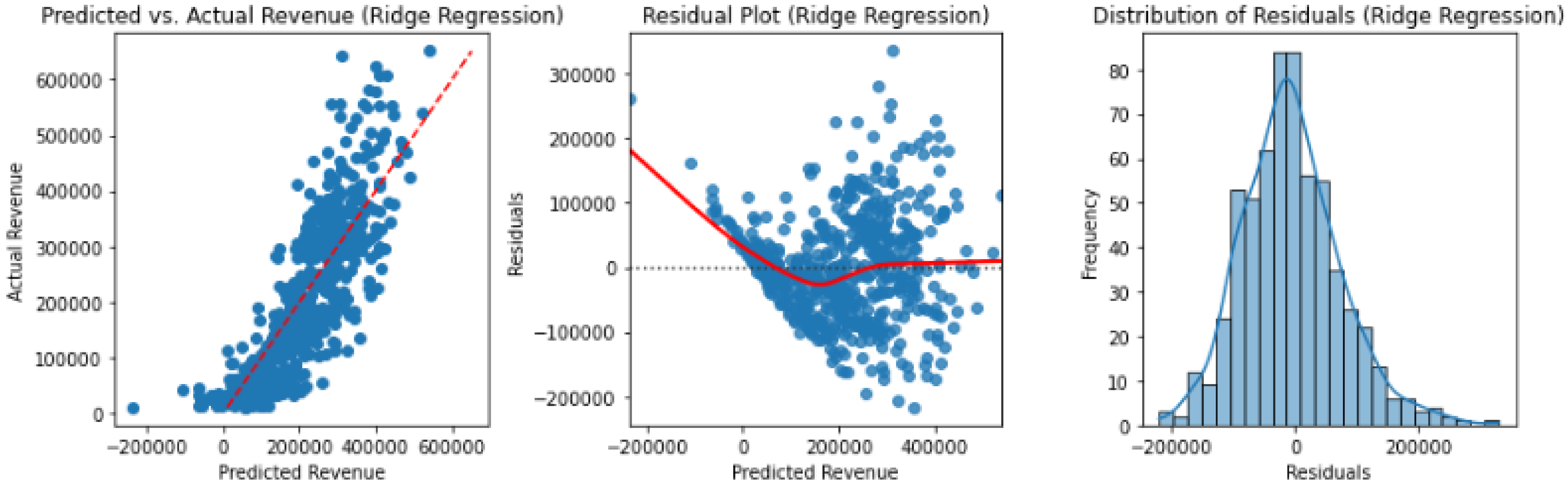


	Model	Train RMSE	Test RMSE	Train MAE	Test MAE	Train R^2	Test R^2
0	Linear Regression	83545.6686	79516.2759	61888.486	61672.2107	0.6873	0.6933

LASSO



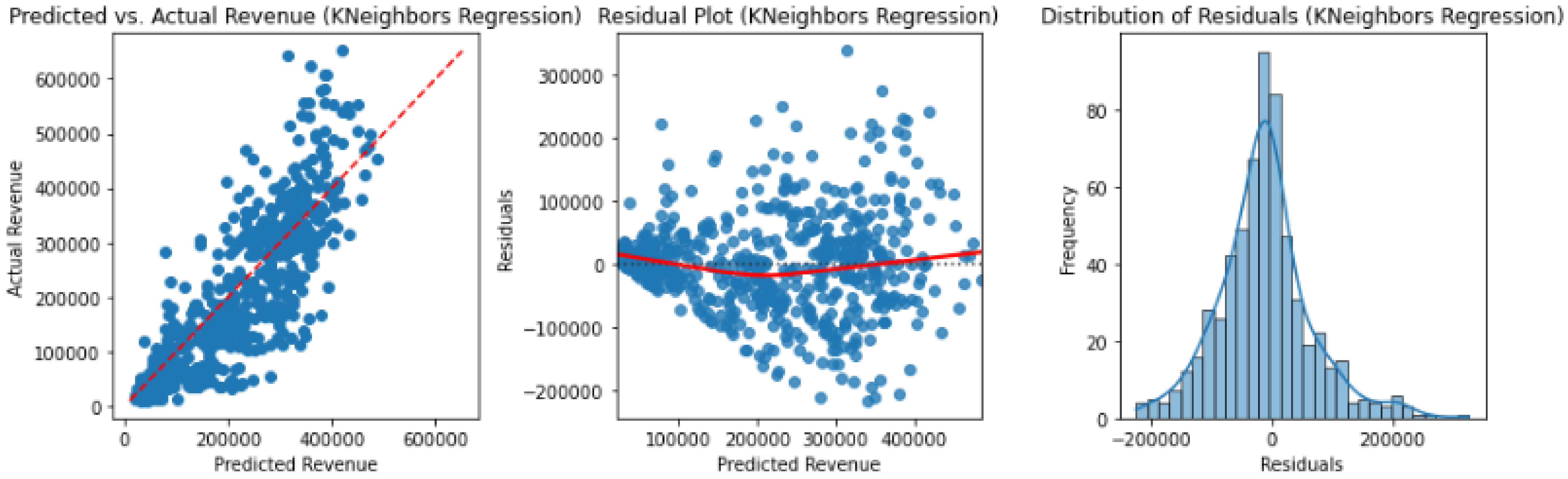
	Model	Train RMSE	Test RMSE	Train MAE	Test MAE	Train R^2	Test R^2
0	Lasso Regression	83545.8857	79513.7191	61891.9644	61667.7066	0.6873	0.6934



RIDGE

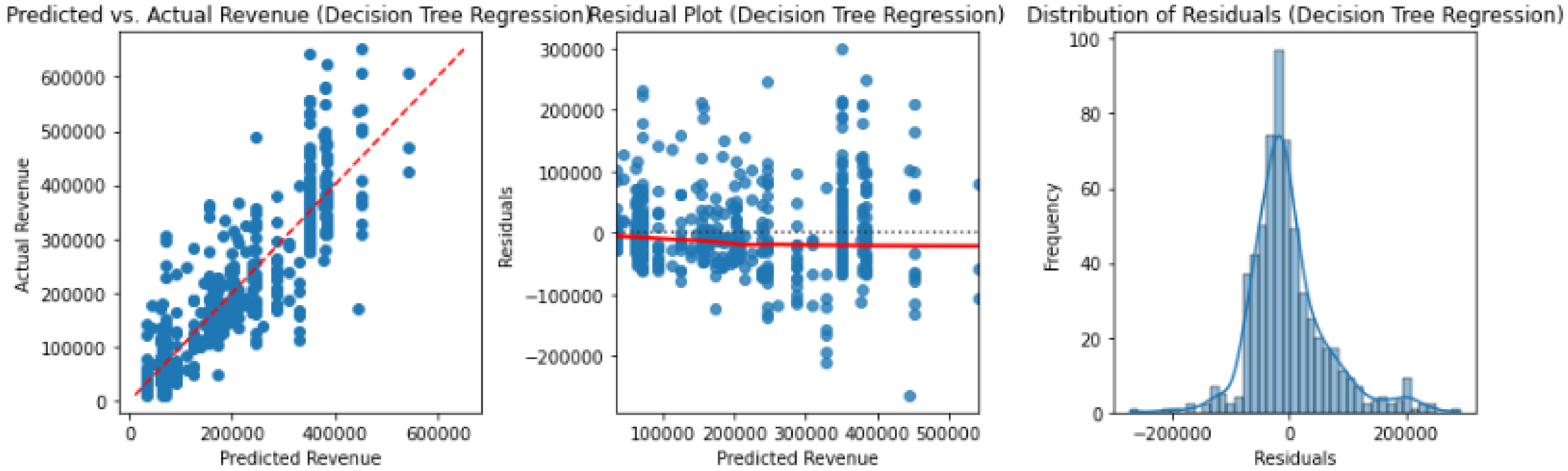
	Model	Train RMSE	Test RMSE	Train MAE	Test MAE	Train R^2	Test R^2
0	Ridge Regression	83686.9855	79608.6407	62062.8609	61602.7549	0.6863	0.6926

KNN



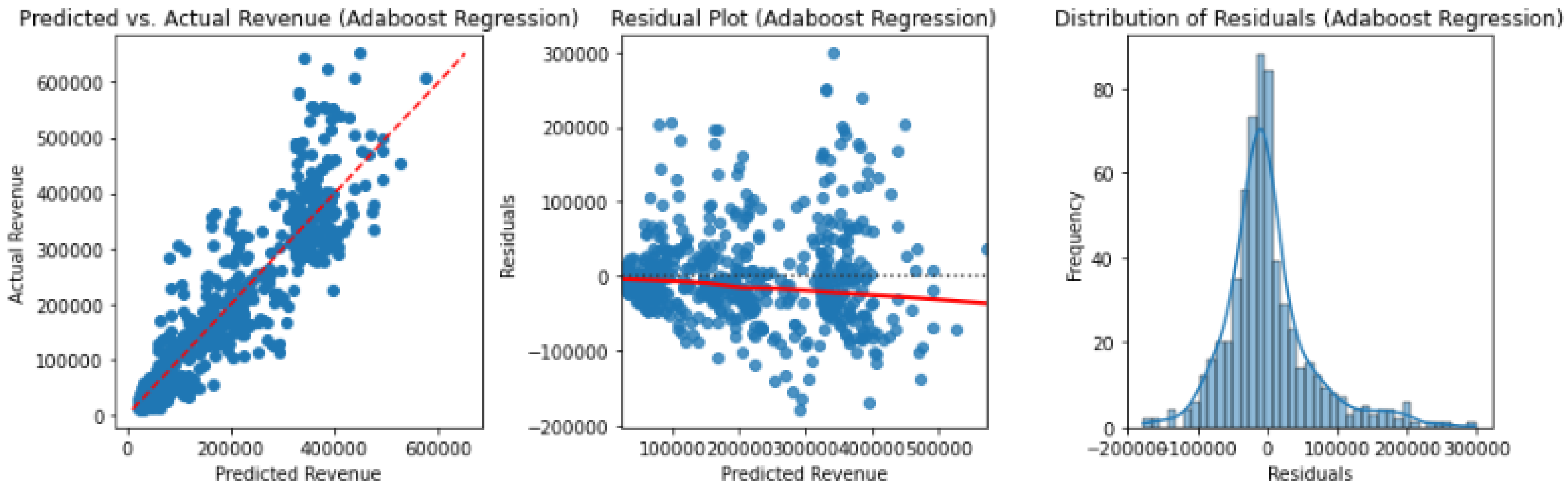
	Model	Train RMSE	Test RMSE	Train MAE	Test MAE	Train R^2	Test R^2
0	KNeighbors Regression	474.8732	78556.6315	21.9353	57644.6915	1.0	0.7007

Decision Tree

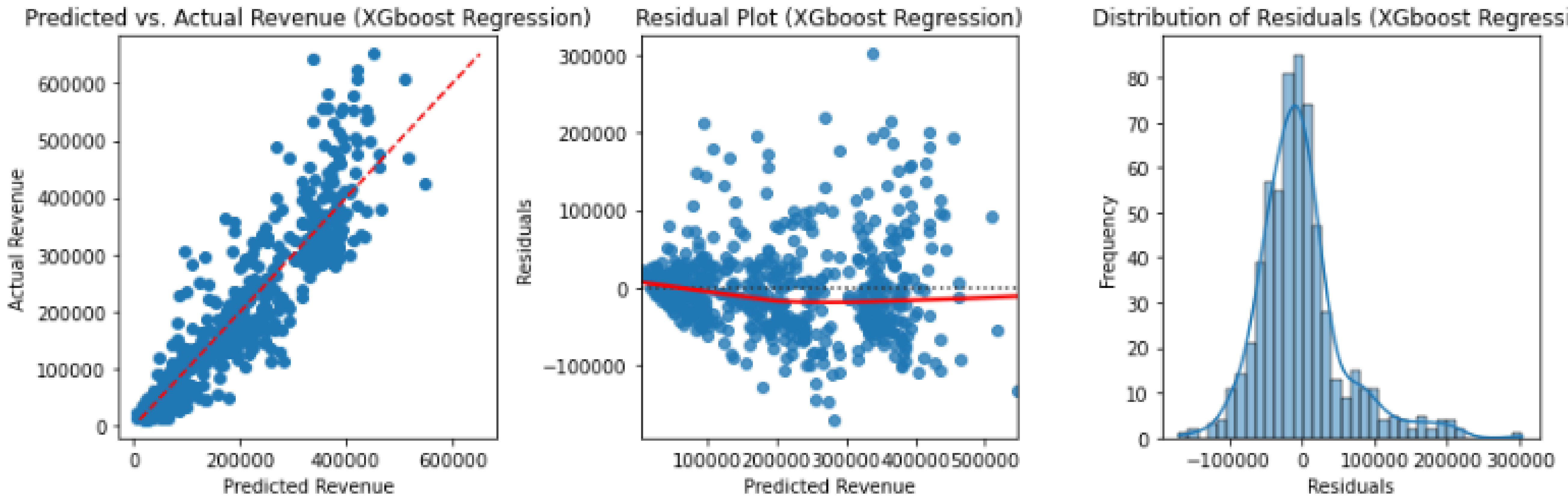


	Model	Train RMSE	Test RMSE	Train MAE	Test MAE	Train R^2	Test R^2
0	Decision Tree Regression	67132.2759	66327.0158	46220.7472	47110.5129	0.7981	0.7866

AdaBoost



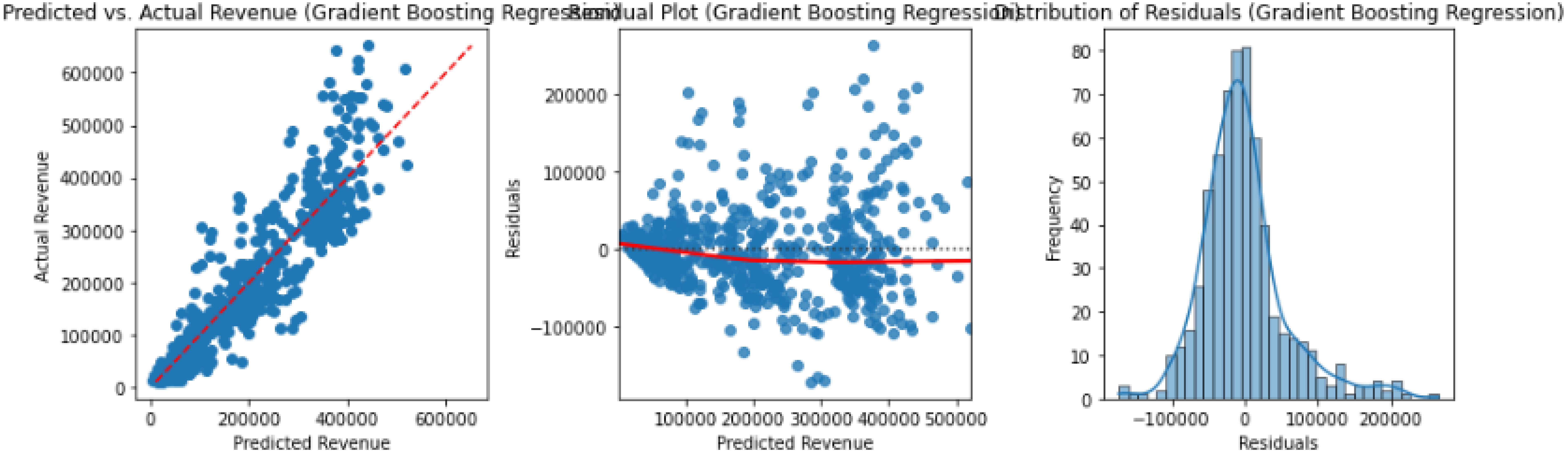
	Model	Train RMSE	Test RMSE	Train MAE	Test MAE	Train R^2	Test R^2
0	Adaboost Regression	28331.4369	62258.2139	19592.183	42463.9576	0.964	0.812



XGBoost

	Model	Train RMSE	Test RMSE	Train MAE	Test MAE	Train R^2	Test R^2
0	XGboost Regression	58242.4314	58894.794	39217.4357	41648.1486	0.848	0.8318

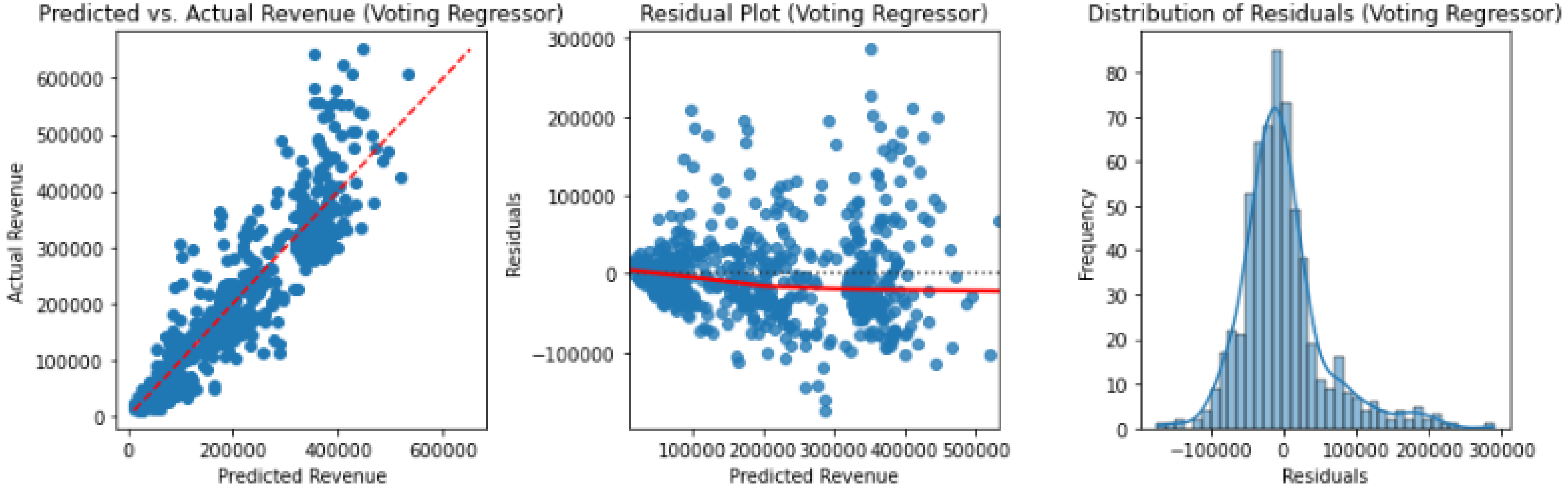
Gradient Boosting Regressor



	Model	Train RMSE	Test RMSE	Train MAE	Test MAE	Train R^2	Test R^2
0	Gradient Boosting Regression	55043.5355	57729.7362	36787.496	40975.2036	0.8643	0.8384

Stacking

	Model	Train RMSE	Test RMSE	Train MAE	Test MAE	Train R^2	Test R^2
0	Stacking Regressor	43716.6124	58025.8784	30743.3438	40870.4684	0.9144	0.8367



Voting

	Model	Train RMSE	Test RMSE	Train MAE	Test MAE	Train R^2	Test R^2
0	Voting Regressor	44715.1882	58243.1908	31333.5127	40769.5361	0.9104	0.8355

Summary

	Model	Train RMSE	Test RMSE	Train MAE	Test MAE	Train R^2	Test R^2
7	Gradient Boosting Regression	55043.5355	57729.7362	36787.4960	40975.2036	0.8643	0.8384
10	Stacking Regressor	43716.6124	58025.8784	30743.3438	40870.4684	0.9144	0.8367
11	Voting Regressor	44715.1882	58243.1908	31333.5127	40769.5361	0.9104	0.8355
6	XGboost Regression	58242.4314	58894.7940	39217.4357	41648.1486	0.8480	0.8318
5	Adaboost Regression	28331.4369	62258.2139	19592.1830	42463.9576	0.9640	0.8120
4	Decision Tree Regression	67132.2759	66327.0158	46220.7472	47110.5129	0.7981	0.7866
3	KNeighbors Regression	474.8732	78556.6315	21.9353	57644.6915	1.0000	0.7007
2	Lasso Regression	83545.8857	79513.7191	61891.9644	61667.7066	0.6873	0.6934
0	Linear Regression	83545.6686	79516.2759	61888.4860	61672.2107	0.6873	0.6933
1	Ridge Regression	83686.9855	79608.6407	62062.8609	61602.7549	0.6863	0.6926
8	Elastic Net Regression	84571.7867	80862.9172	62955.5235	62401.8211	0.6796	0.6829
9	Support Vector Regression	119957.2090	112391.2238	91694.3911	88257.2258	0.3554	0.3873

CONCLUSION/NEXT STEPS

- Summary:
  - the Gradient Boosting Regression stands out as the top performer with the highest test R^2 value (0.8384) and the lowest test RMSE and MAE values among all models.

- Next Steps:
  - Feature Engineering
  - Time-Series Analysis
  - Customer Segmentation

Contact

- David Johnson : Johnsondavidbjr@gmail.com