Predicting Prices of Airbnb Listings

Presented by Qilei Cai and Tasnim Nehal

Dataset

- Obtained from Kaggle.com
- Contains 48895 Airbnb listings located in New York City
- Features include:
  - Neighborhood group (Manhattan, Brooklyn, etc.)
  - Latitude & longitude of accommodation
  - Minimum number of nights required for a booking
  - Total number of reviews received by the host
  - Price
- Goal: train models to predict the price of an Airbnb listing

Unsupervised Analysis

- Graph of longitude vs. latitude

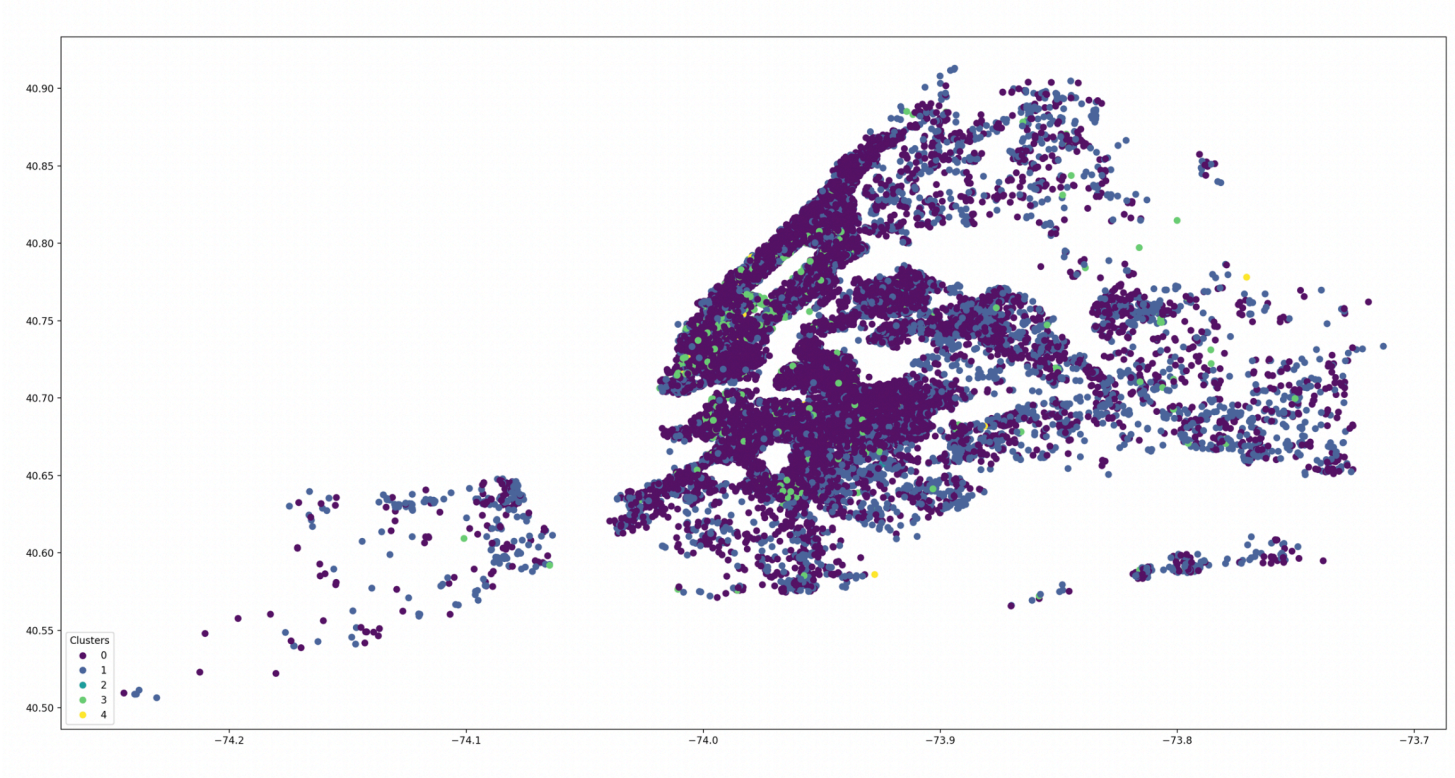


Figure 1: Longitude vs. Latitude (Original)

Unsupervised Analysis

- Graph of longitude vs. latitude

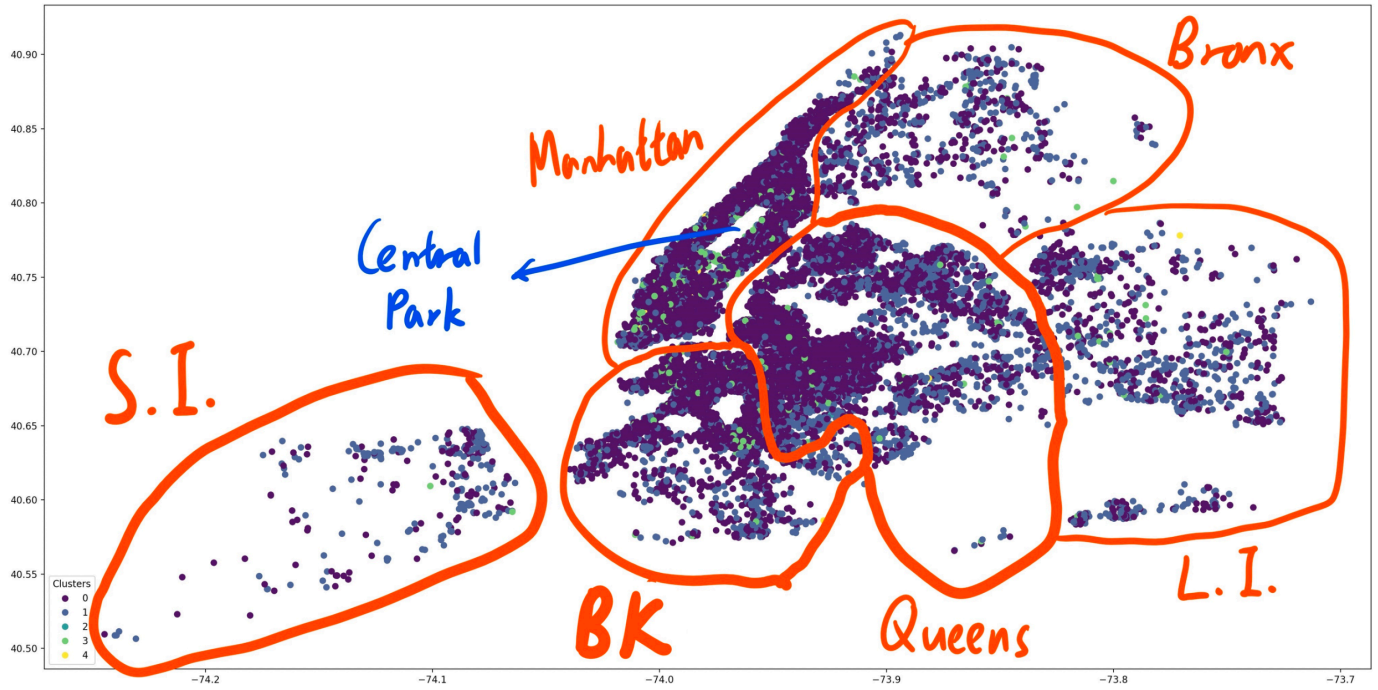


Figure 2: Longitude vs. Latitude (Annotated)

Unsupervised Analysis

- Graph of minimum number of nights vs. number of reviews

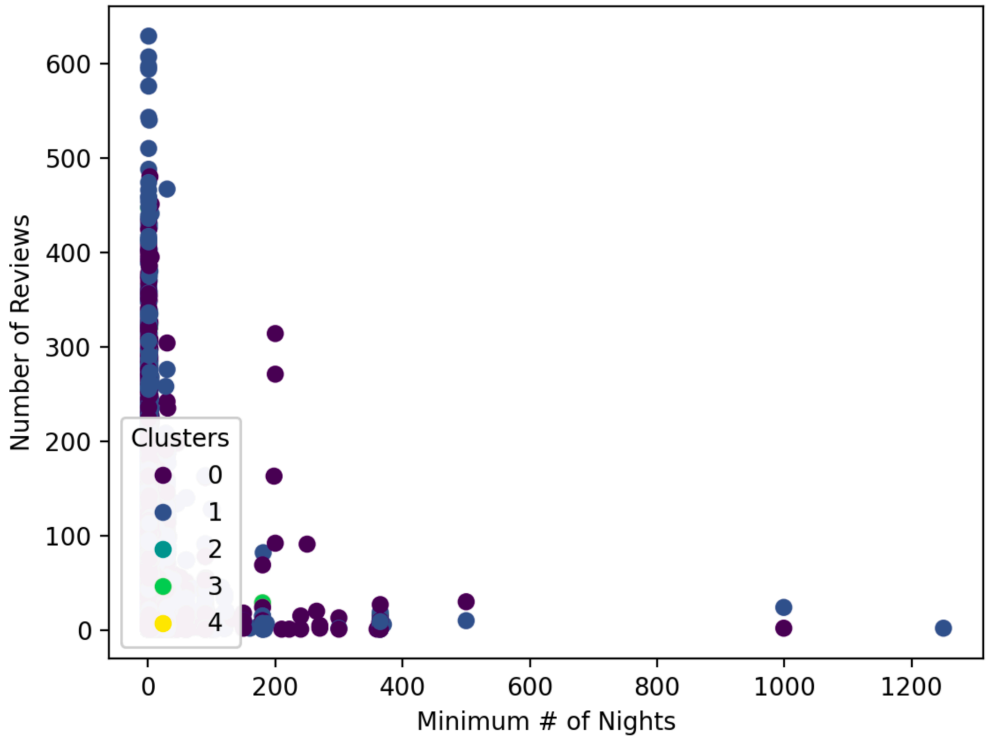


Figure 3: Minimum Number of Nights vs. Number of Reviews

Unsupervised Analysis

- Graph of number of reviews per month vs. price

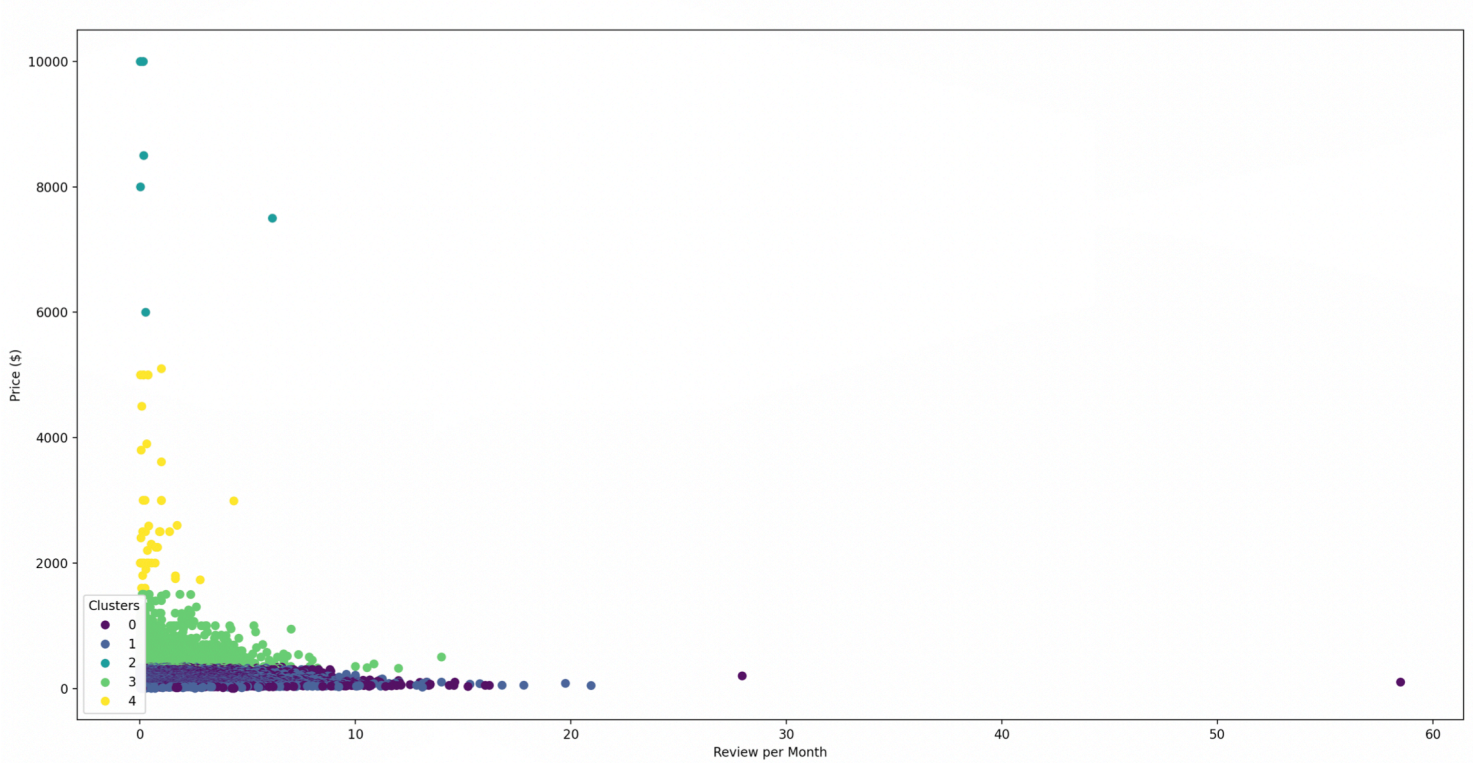


Figure 5: Number of Reviews per Month vs. Price ($)

Unsupervised Analysis

- Graph of room type vs. price

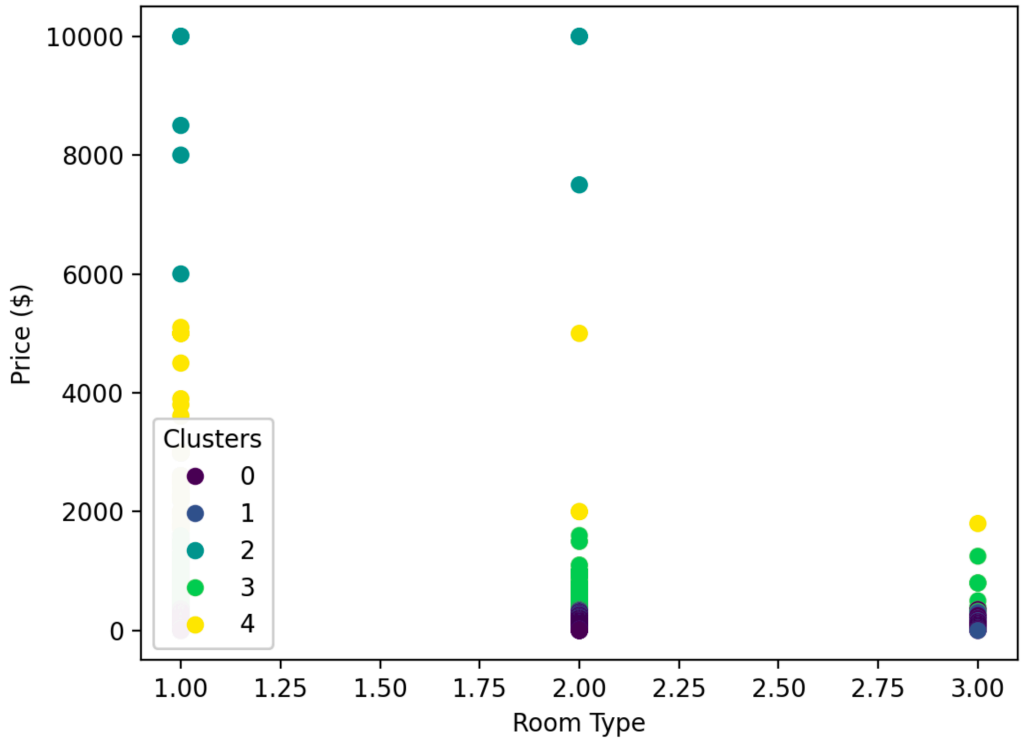


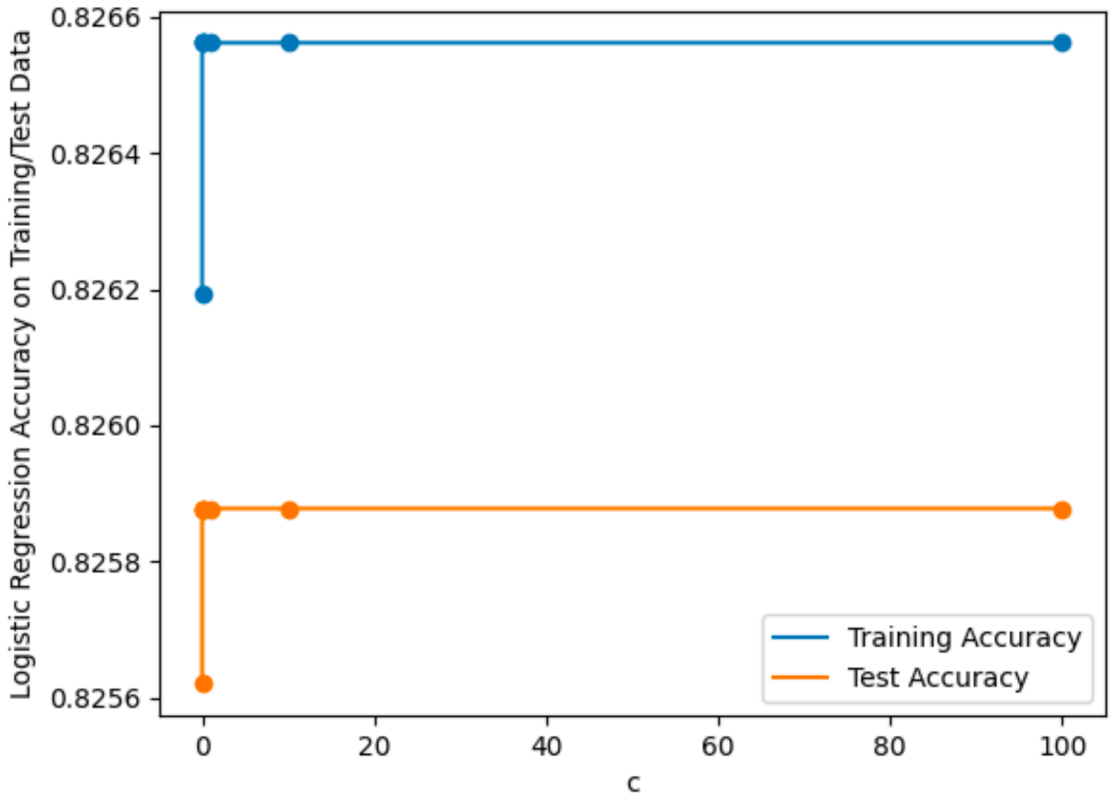
Figure 6: Room Type vs. Price ($)

Supervised Analysis

- Eight price ranges were specified
- Prices were converted to integers from 0 to 7 in accordance
- Label 0: $0-$199.99
- Label 1: $200-$499.99
- Label 2: $500-$999.99
- Label 3: $1000-$1499.99
- Label 4: $1500-$1999.99
- Label 5: $2000-$2999.99
- Label 6: $3000-$3999.99
- Label 7: $4000 and above

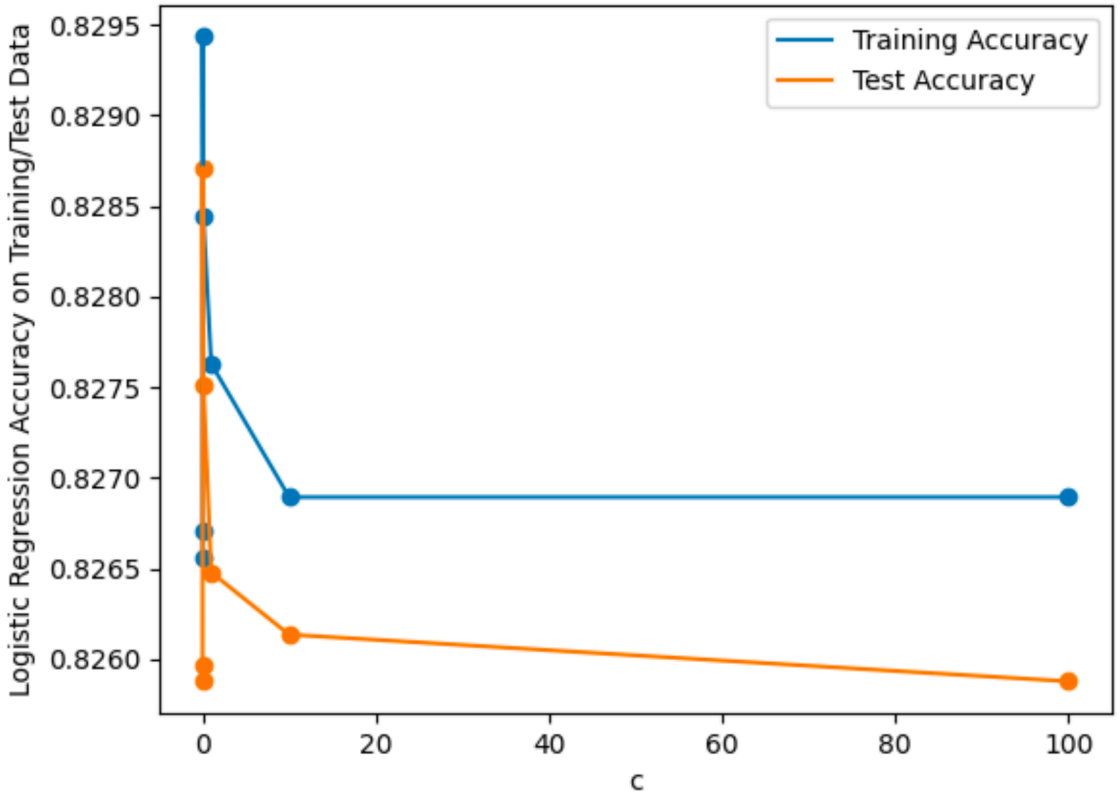
Supervised Analysis

- Logistic regression with L1 regularization



Supervised Analysis

- Logistic regression with L2 regularization



Tables of Results

- Logistic regression

(a) Model with L1 regularization and polynomial feature transformation

	Training Accuracy	Test Accuracy
c=0.0001	0.826525	0.825878
c=0.001	0.826489	0.825878
c=0.01	0.826489	0.825878
c=0.1	0.826489	0.825878
c=1	0.826489	0.825878
c=10	0.826489	0.825878
c=100	0.826525	0.825878

(b) Model with L2 regularization and polynomial feature transformation

	Training Accuracy	Test Accuracy
c=0.0001	0.826525	0.825792
c=0.001	0.825863	0.825277
c=0.01	0.826157	0.825706
c=0.1	0.825789	0.825191
c=1	0.825826	0.825191
c=10	0.825716	0.825191
c=100	0.826341	0.825706

(c) Model with L1 regularization and MinMaxScaler feature transformation

	Training Accuracy	Test Accuracy
c=0.0001	0.826525	0.825878
c=0.001	0.826489	0.825878
c=0.01	0.826489	0.825878
c=0.1	0.826489	0.825878
c=1	0.826489	0.825878
c=10	0.826489	0.825878
c=100	0.826489	0.825878

(d) Model with L2 regularization and MinMaxScaler feature transformation

	Training Accuracy	Test Accuracy
c=0.0001	0.826525	0.825792
c=0.001	0.825863	0.825277
c=0.01	0.826157	0.825706
c=0.1	0.825789	0.825191
c=1	0.825826	0.825191
c=10	0.825716	0.825191
c=100	0.826341	0.825706

Tables of Results

- Support Vector Machines

(a) Model with linear kernel

	Training Accuracy	Test Accuracy
c=0.0001	0.815951	0.811429
c=0.001	0.815951	0.811429
c=0.01	0.815951	0.811429
c=0.1	0.815951	0.811429
c=1	0.815951	0.811429
c=10	0.815951	0.811429
c=100	0.815951	0.811429

(b) Model with radial-basis function kernel

	Training Accuracy	Test Accuracy
c=0.0001	0.823313	0.820000
c=0.001	0.823313	0.820000
c=0.01	0.823313	0.820000
c=0.1	0.823313	0.820000
c=1	0.823313	0.820000
c=10	0.823313	0.820000
c=100	0.823313	0.820000

(c) Model with polynomial kernel

	Training Accuracy	Test Accuracy
c=0.0001	0.823313	0.820000
c=0.001	0.823313	0.820000
c=0.01	0.823313	0.820000
c=0.1	0.823313	0.820000
c=1	0.823313	0.820000
c=10	0.823313	0.820000
c=100	0.824540	0.820000

Tables of Results

- Neural Networks

(a) Model with L1 regularization and polynomial feature transformation

	Training Accuracy	Test Accuracy
c=0.0001	0.541104	0.551428
c=0.001	0.503067	0.528571
c=0.01	0.477300	0.491428
c=0.1	0.796319	0.814285
c=1	0.802453	0.817142
c=10	0.765644	0.791428
c=100	0.396319	0.445714

(b) Model with L2 regularization and polynomial feature transformation

	Training Accuracy	Test Accuracy
c=0.0001	0.802453	0.817142
c=0.001	0.801226	0.817142
c=0.01	0.785276	0.800000
c=0.1	0.798773	0.811428
c=1	0.456441	0.474285
c=10	0.802453	0.817142
c=100	0.802453	0.817142

(c) Model with L1 regularization and MinMaxScaler feature transformation

	Training Accuracy	Test Accuracy
c=0.0001	0.008588	0.005714
c=0.001	0.000000	0.005714
c=0.01	0.496932	0.482857
c=0.1	0.096932	0.140000
c=1	0.013496	0.008571
c=10	0.013496	0.005714
c=100	0.348466	0.317142

(d) Model with L2 regularization and MinMaxScaler feature transformation

	Training Accuracy	Test Accuracy
c=0.0001	0.495705	0.497142
c=0.001	0.014723	0.005714
c=0.01	0.166871	0.165714
c=0.1	0.004907	0.002857
c=1	0.004907	0.002857
c=10	0.000000	0.000000
c=100	0.001226	0.002857

Conclusion

- Highest test accuracy: Logistic regression with L1 regularization and polynomial feature transformation
- Lowest test accuracy on average: neural networks with MinMaxScaler feature transformation
- Neural networks with polynomial feature transformations yielded decent test accuracy
- MinMaxScaler feature transformation yielded decent accuracy on logistic regression models
- Same type of feature transformation can deliver vastly different performance on different types of models

Thank you for listening