Rice Datathon 2021 - Chevron

German Gonzalez

Preprocessing

Dropped

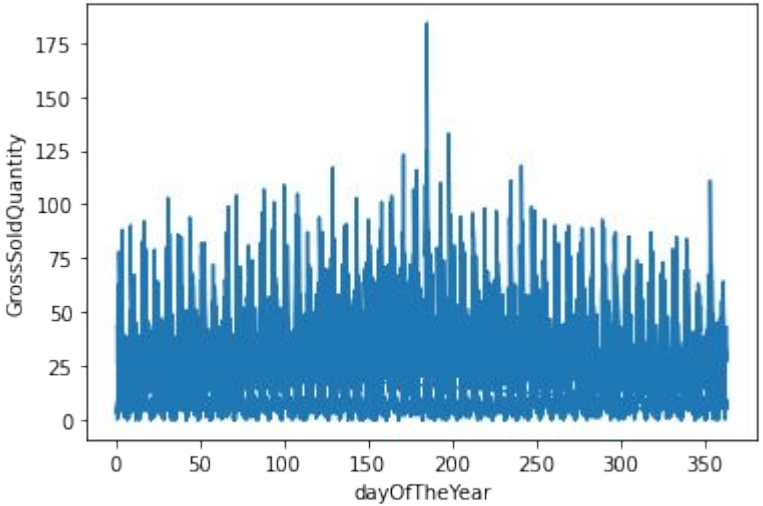
- Cash/Credit Site
- Loyalty Site
- Extra Mile Site
- CoBrand
- Food Service
- State

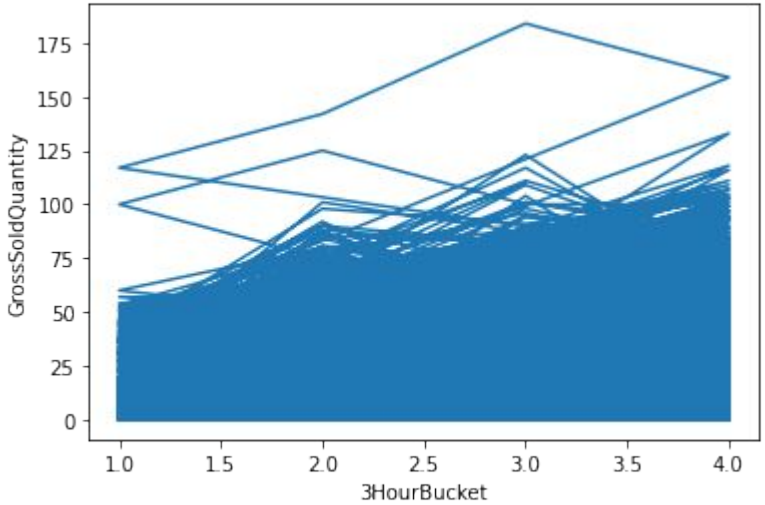
df[item].unique()

Scaled ‘StoreNumber’ by 1000

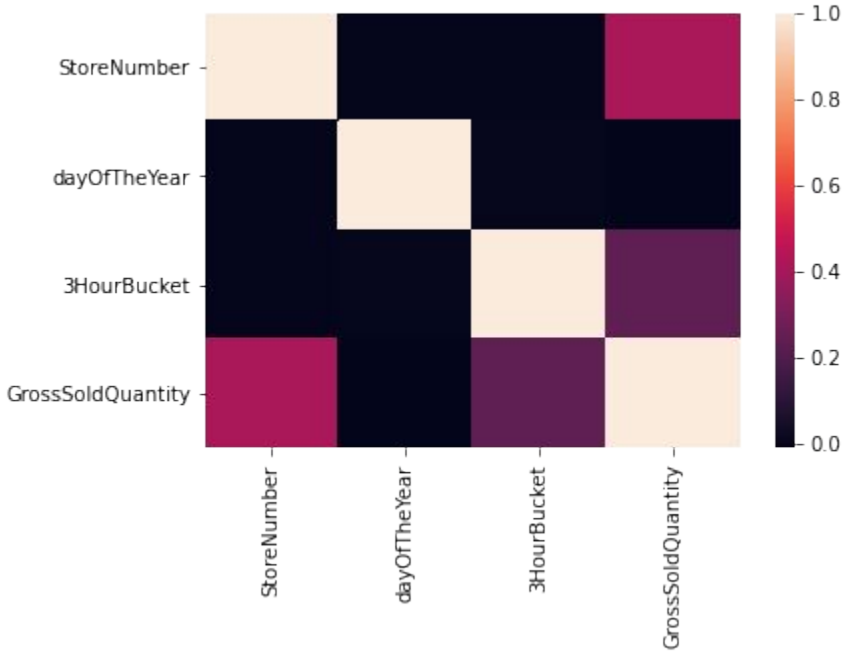
Split the data into Train, Validation, Testing

Visualizations





Quick Correlations Matrix



Feature Representation/Columns

Mapped ‘City’ column as a categorical column

Bucketized ‘Store Number’, ‘3HourBucket’, ‘dayOfTheYear’

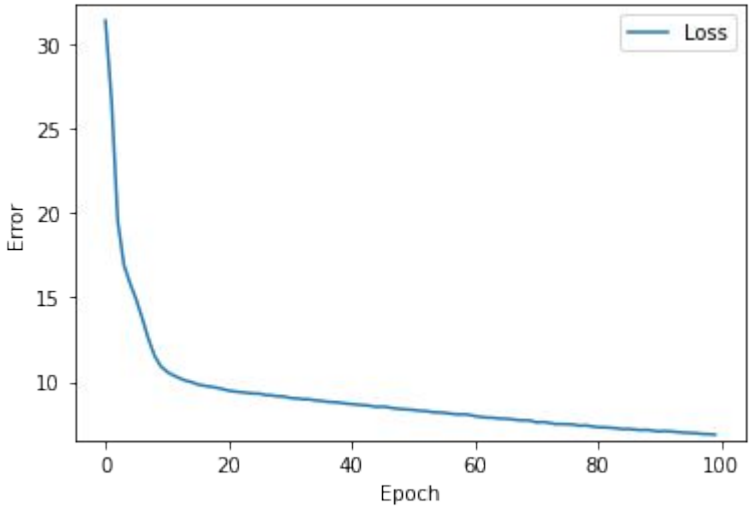
Crossed the following together: ‘EBT Site’, ‘Alcohol’, ‘Carwash’

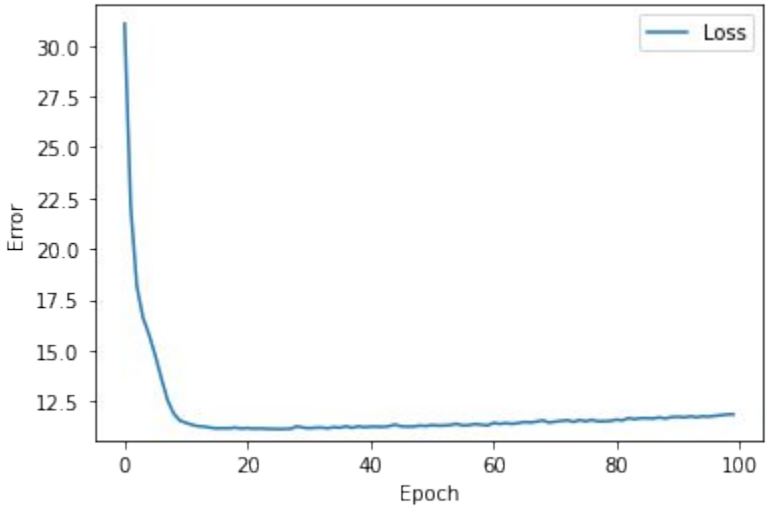
Model

Neural Network

- Feature Layer
- Dense 16 units, relu Layer
- Dense 8 units, relu Layer
- Dense 4 units, relu Layer
- Dense 1 unit, linear Layer

Loss Curves


Loss
Error
Epoch
30
25
20
15
10
0
20
40
60
80
100


Loss
Error
Epoch
30.0
27.5
25.0
22.5
20.0
17.5
15.0
12.5
0
20
40
60
80
100

RMSE Result

NN model RMSE result - 11.285