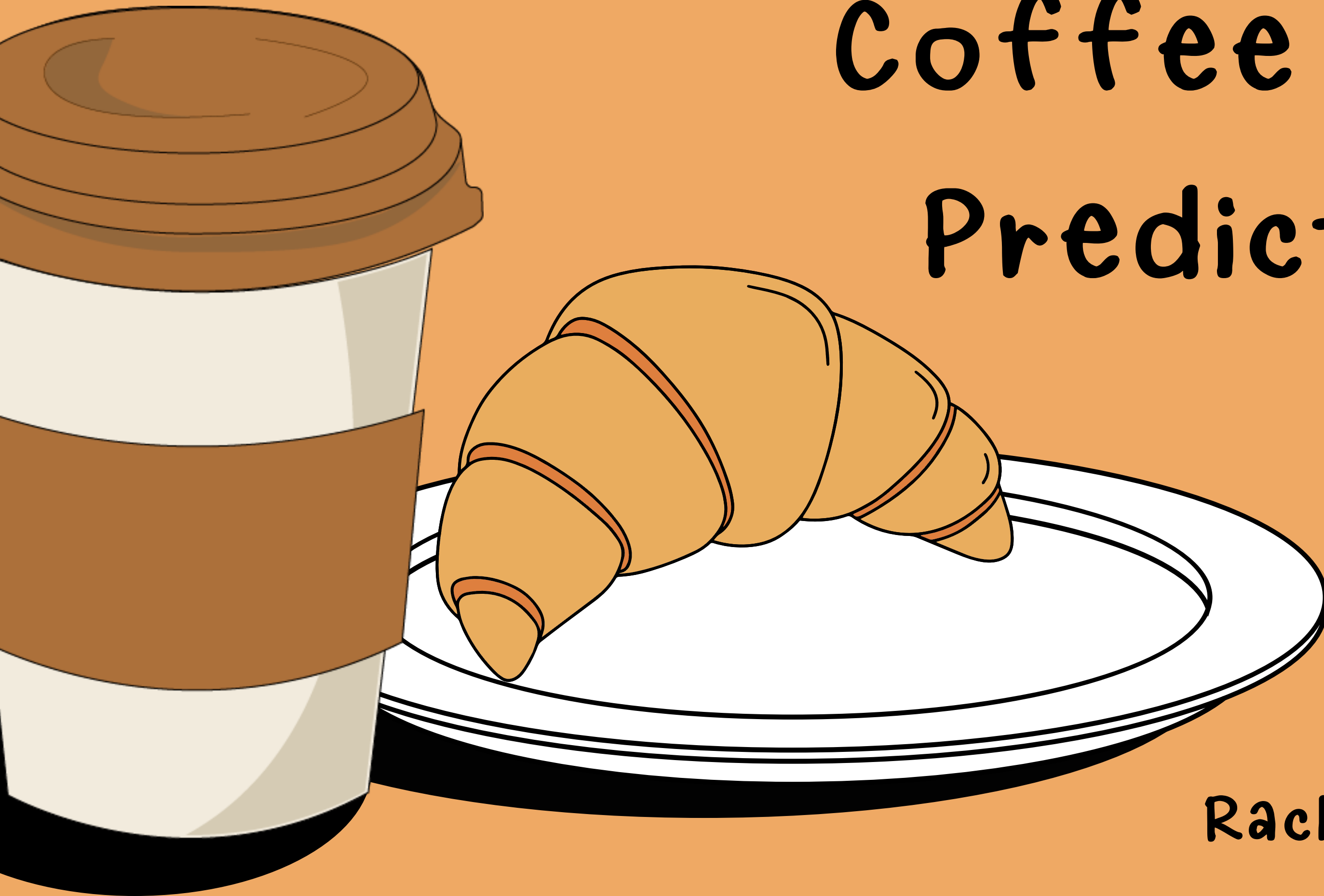


Coffee Sales Prediction



Rachata Kaewviset

Goal

predict quantity of each menu
we can sales each day

sunday



4 cup

monday



6 cup

tuesday



3 cup

Dataset

≈ 150000 transaction from coffee shop from 3 town in usa

Lower Manhattan
33.5%

Astoria
33.1%



Hell's Kitchen
33.4%



feature selection

feature we have

- transaction id
- transaction date
- transaction time
- store id
- store location
- product id
- transaction quantity
- unit price
- Total Bill
- product category
- product type
- product detail
- Size
- Month
- Name
- Day Name
- Hour
- Month
- Day of Week



feature we want

- unit price
- product category
- product type
- Size
- Day Name
- transaction quantity (target value)

data cleaning and feature selection

```
grouped_df = df.groupby(['transaction_date', 'store_location', 'unit_price', 'product_category',  
                        'product_type', 'Size', 'Day Name'])['transaction_qty'].sum().reset_index()
```

*we still keep transaction date and store location here because we want to separate sales of each day and each location

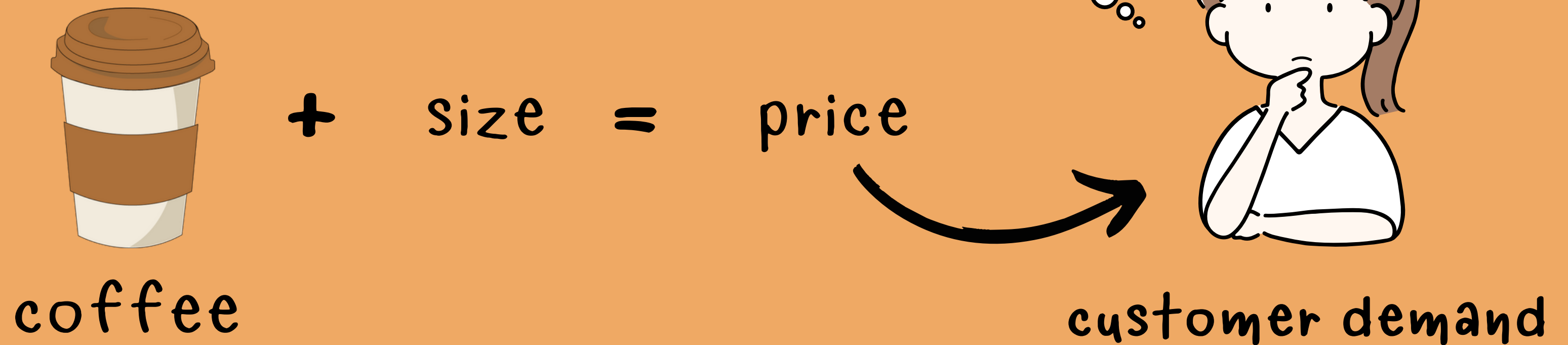


```
df = df.drop(labels: ['transaction_date'], axis=1)  
df = df.drop(labels: ['store_location'], axis=1)
```

we drop it later

Linear Regression

since all data was relate to each other
we can use linear regression to predict
the sales quantity



Absolute loss

We chose absolute loss for predicting coffee shop sales because it gives a clear picture of how many sales we are missing in our predictions

± 3



First test

abs loss ≈ 4.3

Prediction Problem

```
def predict(self, X):  
    X_b = np.c_[np.ones((len(X), 1)), X]  
    return (X_b.dot(np.hstack((self.intercept_, self.coef_))))
```

sometimes it can predict
negative and float value

i would like to
order -3.2589 cup
of coffee

huh?



Prediction Problem

```
def predict(self, X):  
    X_b = np.c_[np.ones((len(X), 1)), X]  
    predictions = X_b.dot(np.hstack((self.intercept_, self.coef_)))  
    predictions[predictions < 0] = 0  
    return np.floor(predictions).astype(int)
```

if it less than 0 its 0 and
round everything down

0 cup of coffee
and 1 cup of tea



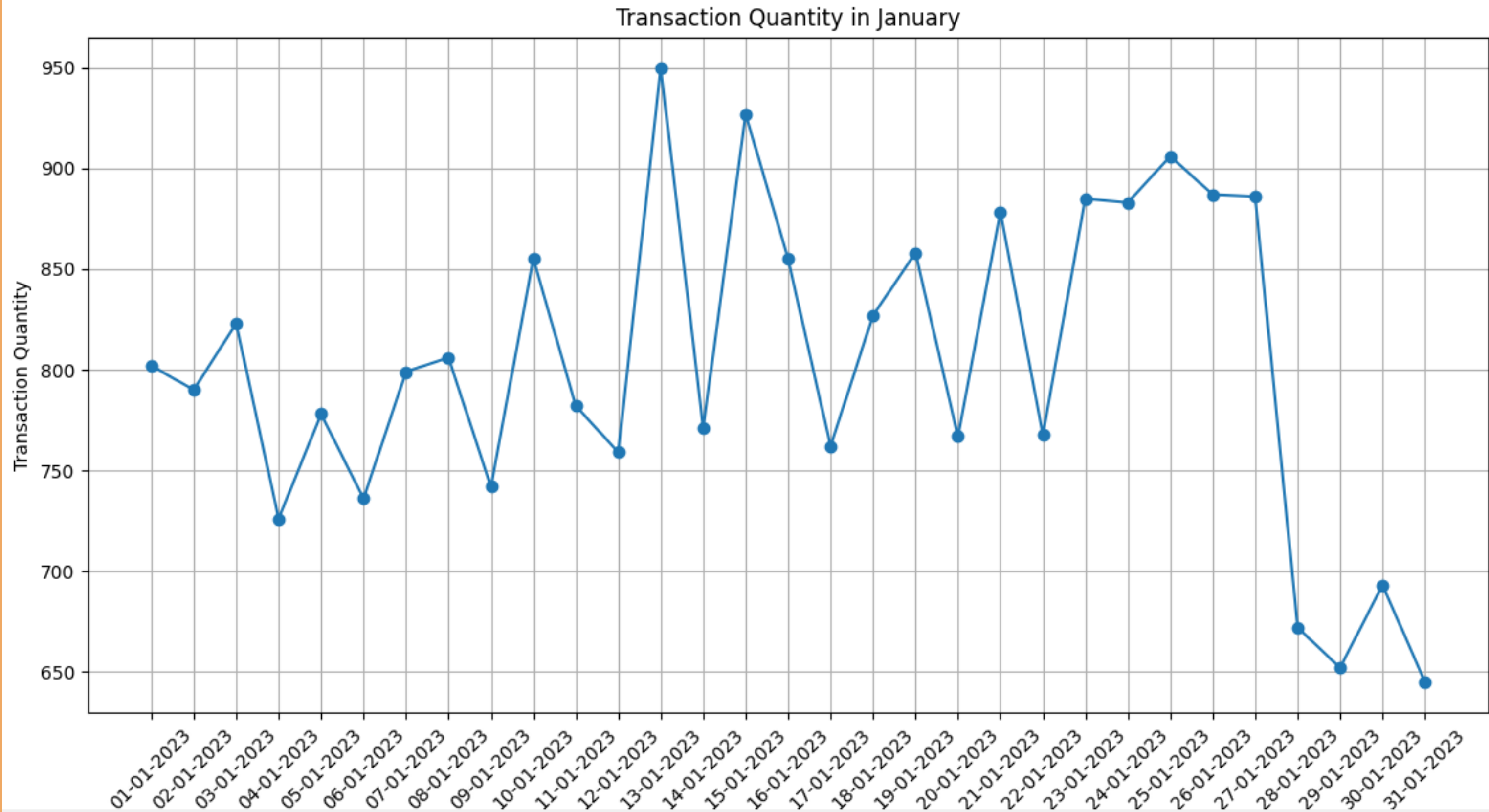
Second test

abs loss ≈ 4.15

What's
problem



Outliers



Normalize

$$x = \frac{x - \min(x)}{x(\max) - \min(x)}$$

```
df['transaction_qty'] = (df['transaction_qty'] - df['transaction_qty'].min()) / (df['transaction_qty'].max() - df['transaction_qty'].min())  
df['unit_price'] = (df['unit_price'] - df['unit_price'].min()) / (df['unit_price'].max() - df['unit_price'].min())
```




final test

abs loss ≈ 0.12