BESPOKE™

AQUARIUMS

Sales Prediction

Bijal Saija

Business Problem

The key business problem identified for stock maintenance and to keep the money flow controlled. Accurate sales forecasts are crucial for effective management in an e-commerce business. This project focuses on predicting sales for an online retail store, specifically Bespoke Aquariums, using data from actual transactions. Accurate sales forecasts enable retailers to more precisely plan for the future of an e-commerce business. Sales forecasting is an important part of meeting customer expectations and provides insight into how the market will react to any given product.

The Data

The data is taken from bespoke_data.csv

This data is the Bespoke Aquariums e-commerce Sales dataset.

Features used for analysis

Order total amount (target variable)	
Order Number (continuous)	Order Number (continuous)
Quantityordered (continuous)	Orderdate (continuous)
Public Holiday (categorical)	Weekend (categorical)
State (categorical)	Shipping Method Title (categorical)

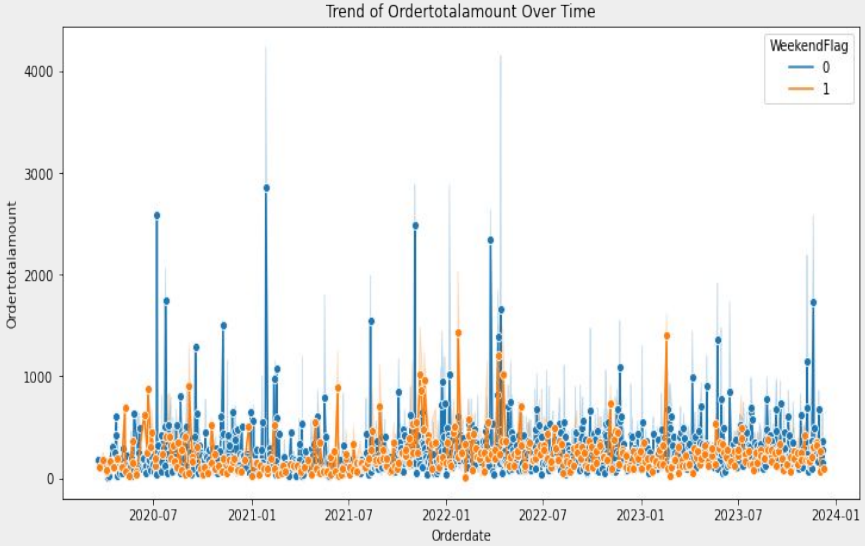
Data used for a feature engineering

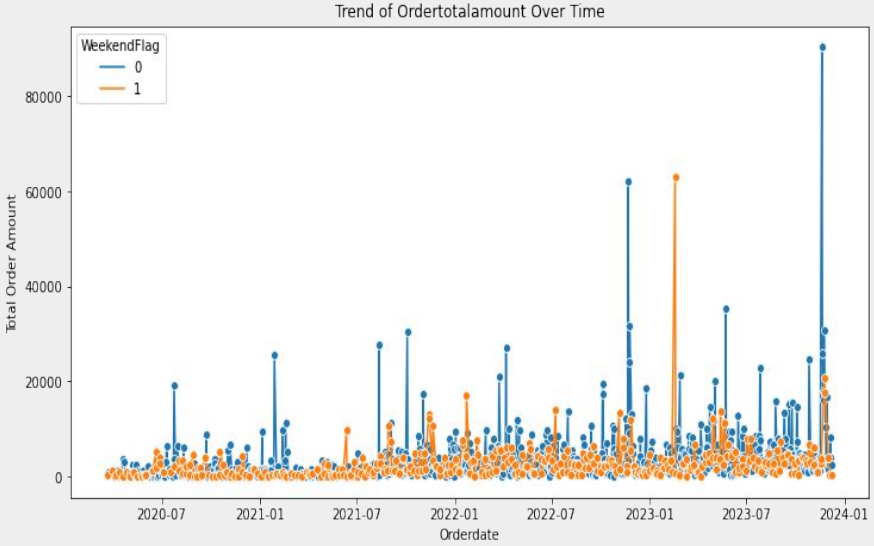
The data is taken from data.gov.au

Feature name is Public Holiday : Merging this data with bespoke data resulted 1 for public holiday otherwise 0.

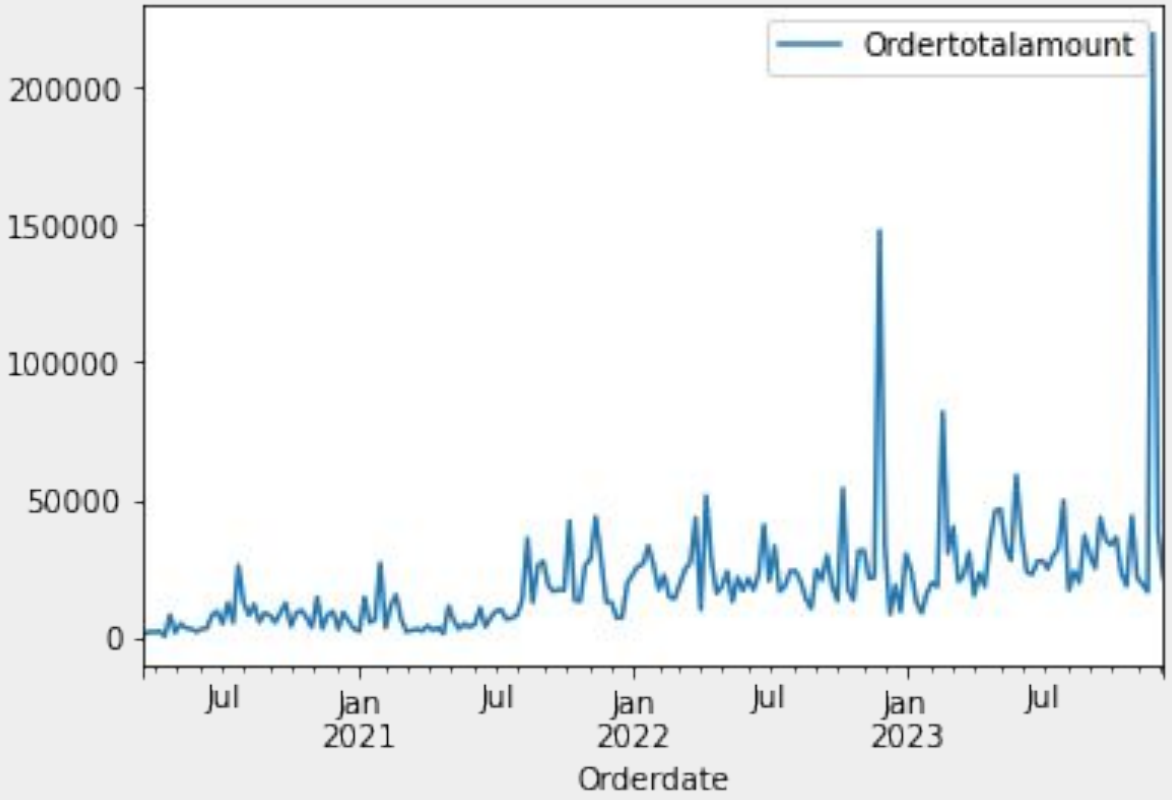
	Order Number	Orderdate	Ordertotalamount	Quantityordered	State	Shipping Method Title	Weekend	Public Holiday
0	95599	2023-12-09	94.5	1	Western Australia	Flat rate	1	0
1	95598	2023-12-09	87.5	1	New South Wales	Flat rate	1	0
2	95597	2023-12-09	89.5	2	Australian Capital Territory	Flat rate	1	0
3	95595	2023-12-09	128.5	1	Queensland	Flat rate	1	0
4	95594	2023-12-09	54.5	1	Victoria	Flat rate	1	0

Order total amount over time before and after building a model

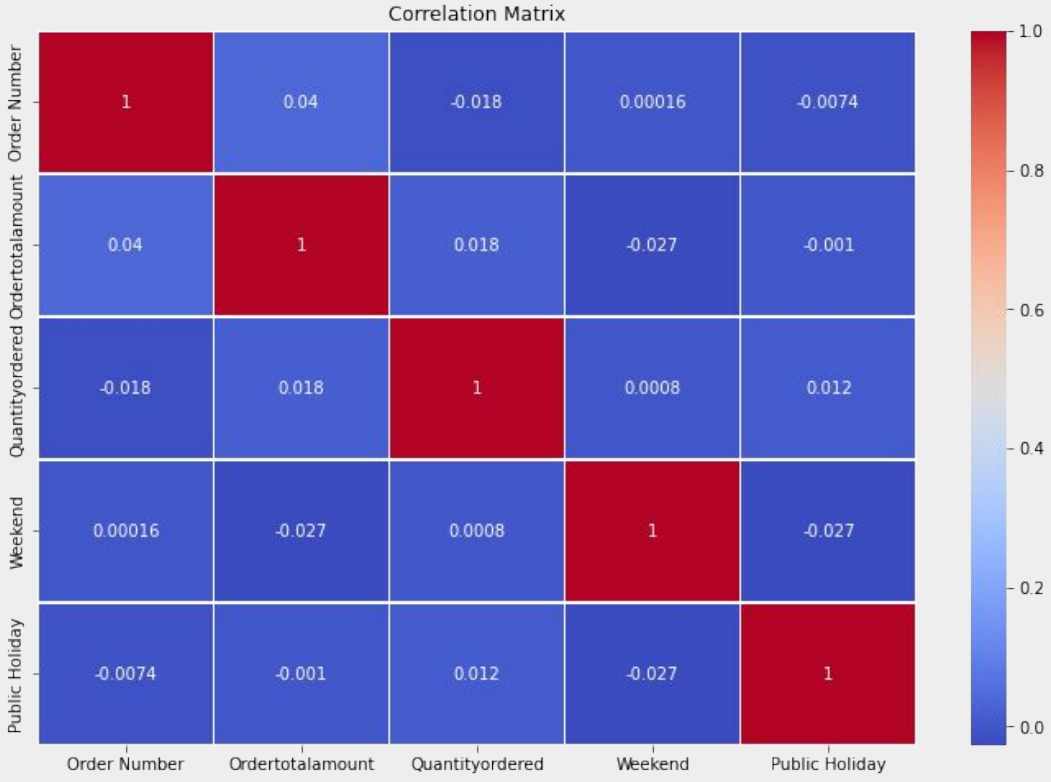




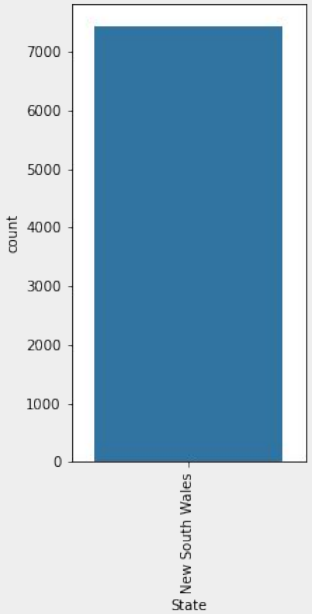
Trend of order total amount over the time

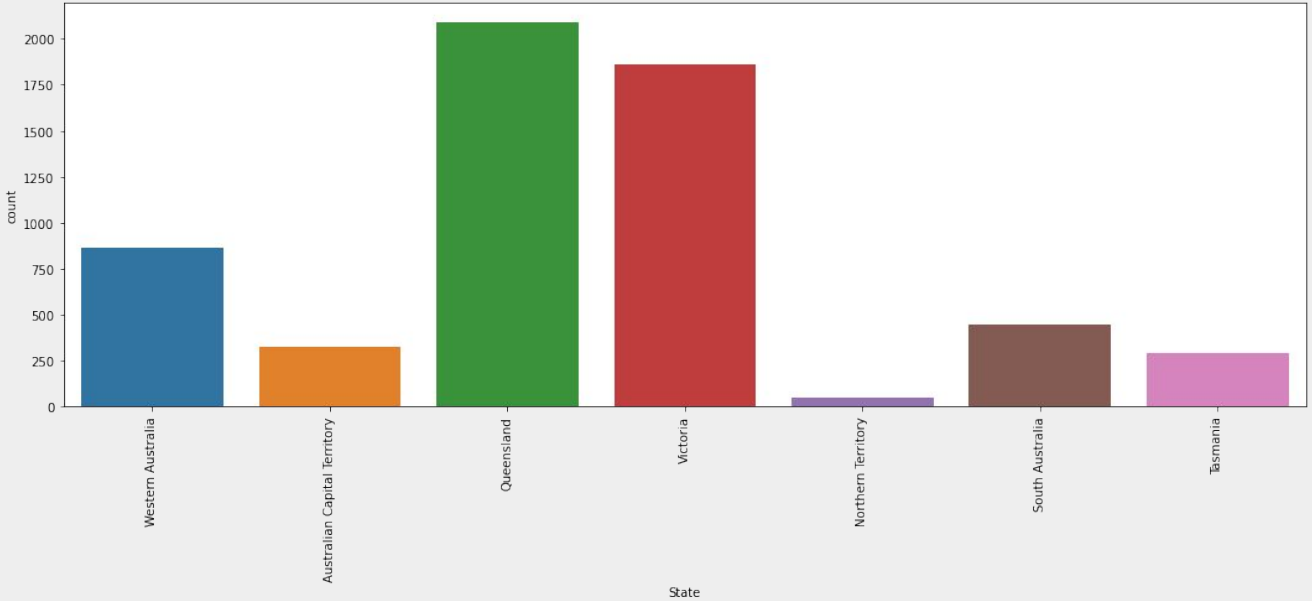


Correlation between continuous variables



Visualizing state vs number of orders to suggest marketing strategies





Building model

Creating dummies

Creating dummies for categorical columns Shipping method title and state

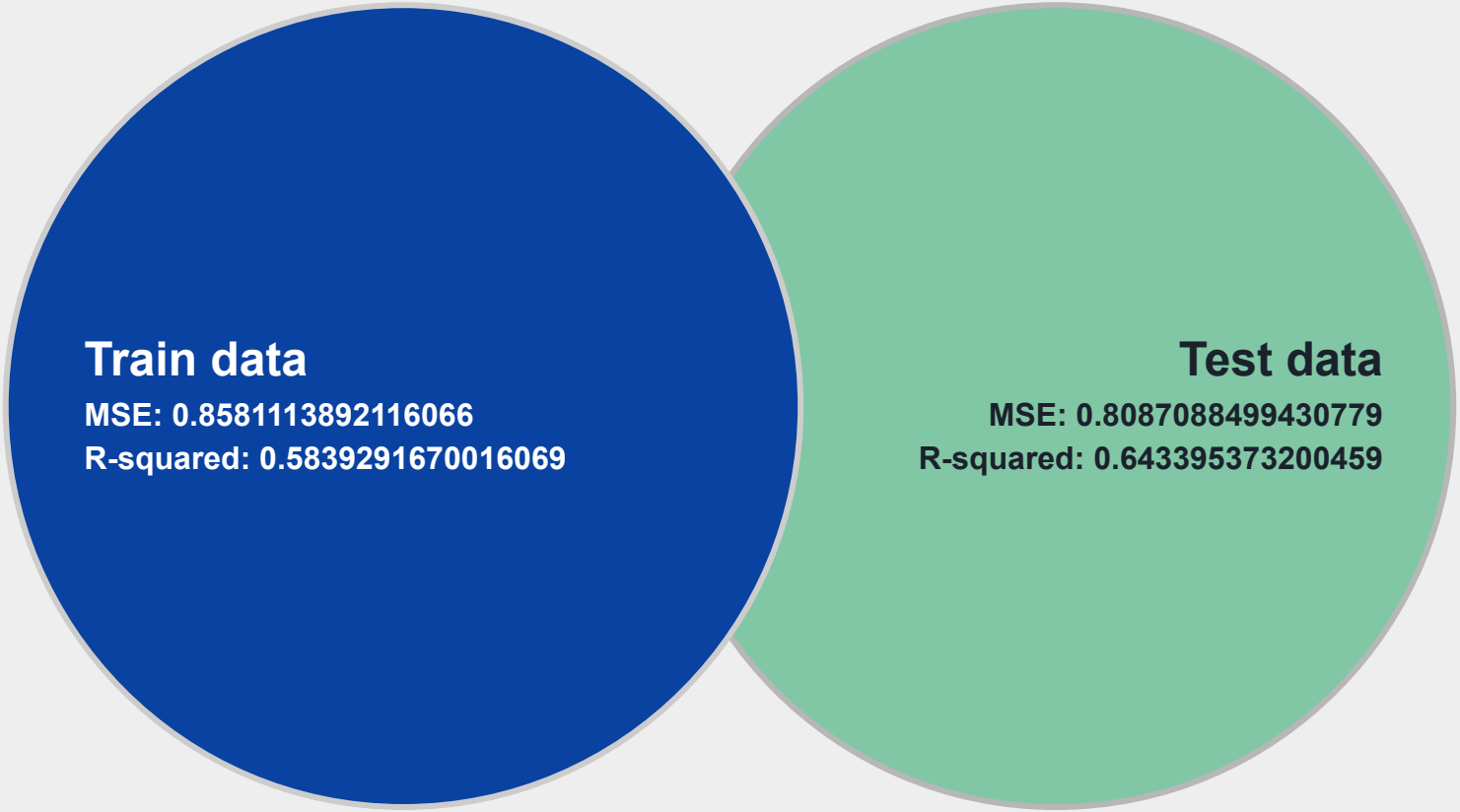
Groupby Orderdate and aggregate

Group by date and calculate total orders, total order amount

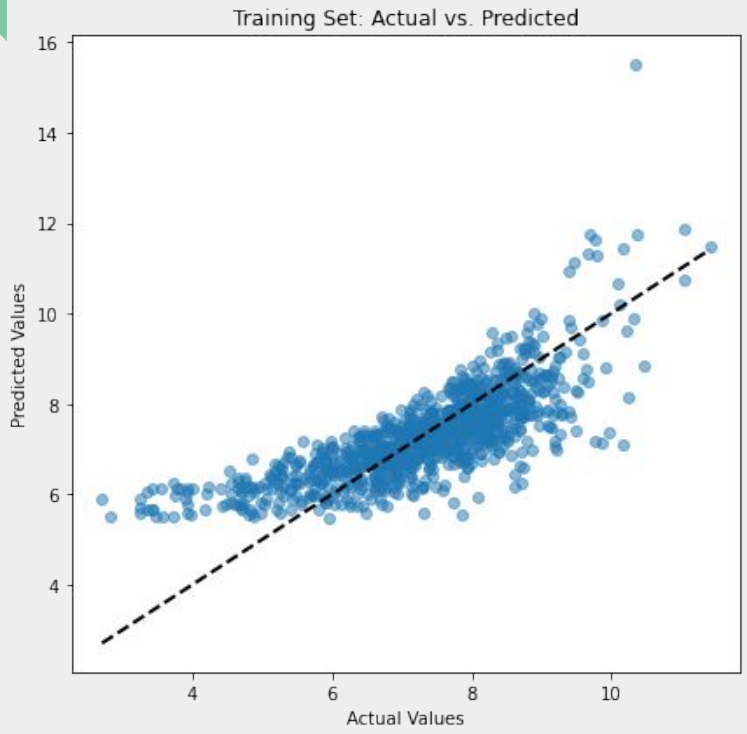
Log transformation of target varible

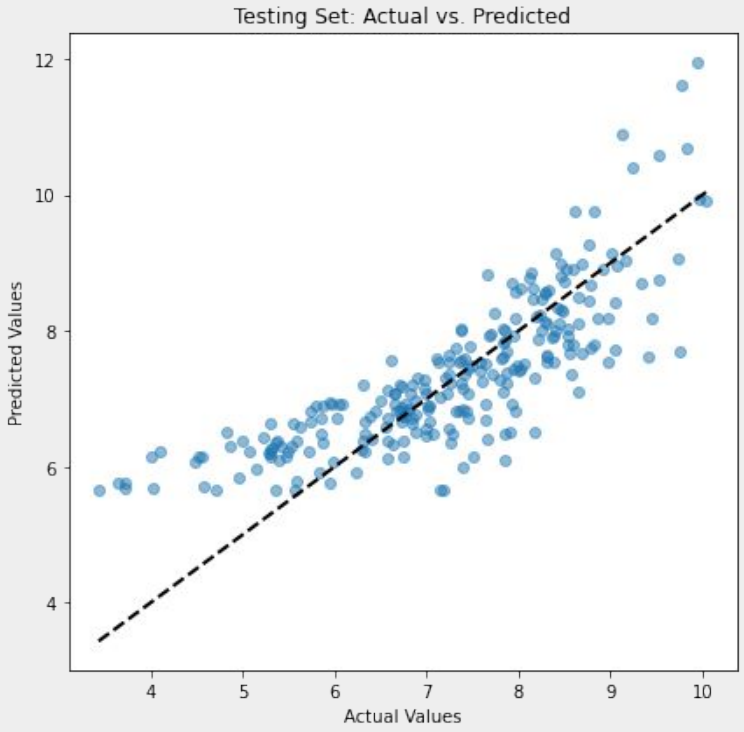
Log-transform the sum of Ordertotalamount

Model Evaluation

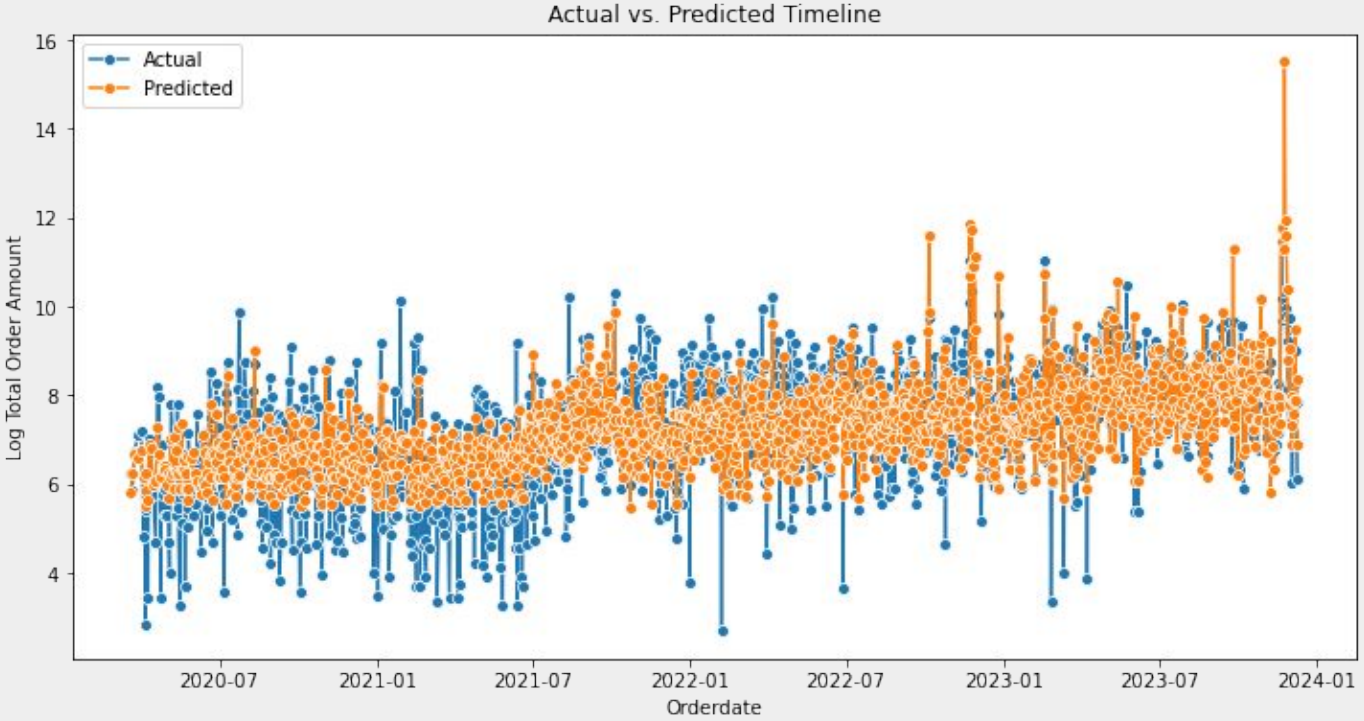


Model Evaluation





Timeline of actual vs predicted data



Conclusion

The most influential feature appears to be "Total Orders" with a coefficient of 0.0864, suggesting that a one-unit increase in total orders is associated with an increase of 0.0864 in the log total order amount.

Some features, such as "State_New South Wales_max," "State_Queensland_max," "State_Victoria_max," "Shipping_Flat rate_max," and "Shipping_Local pickup_max," appear to be statistically significant (p-value < 0.05).

Thank You!