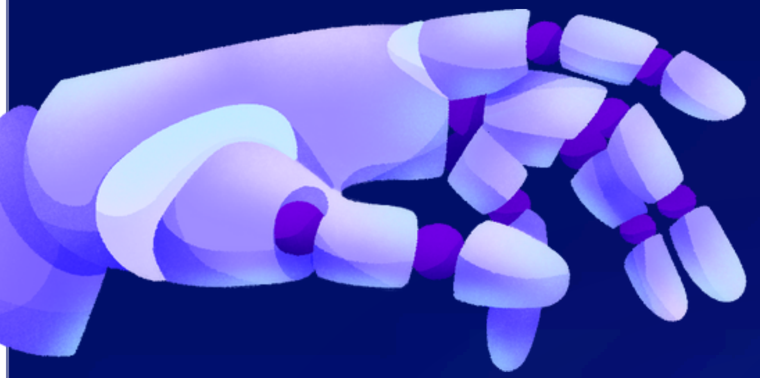
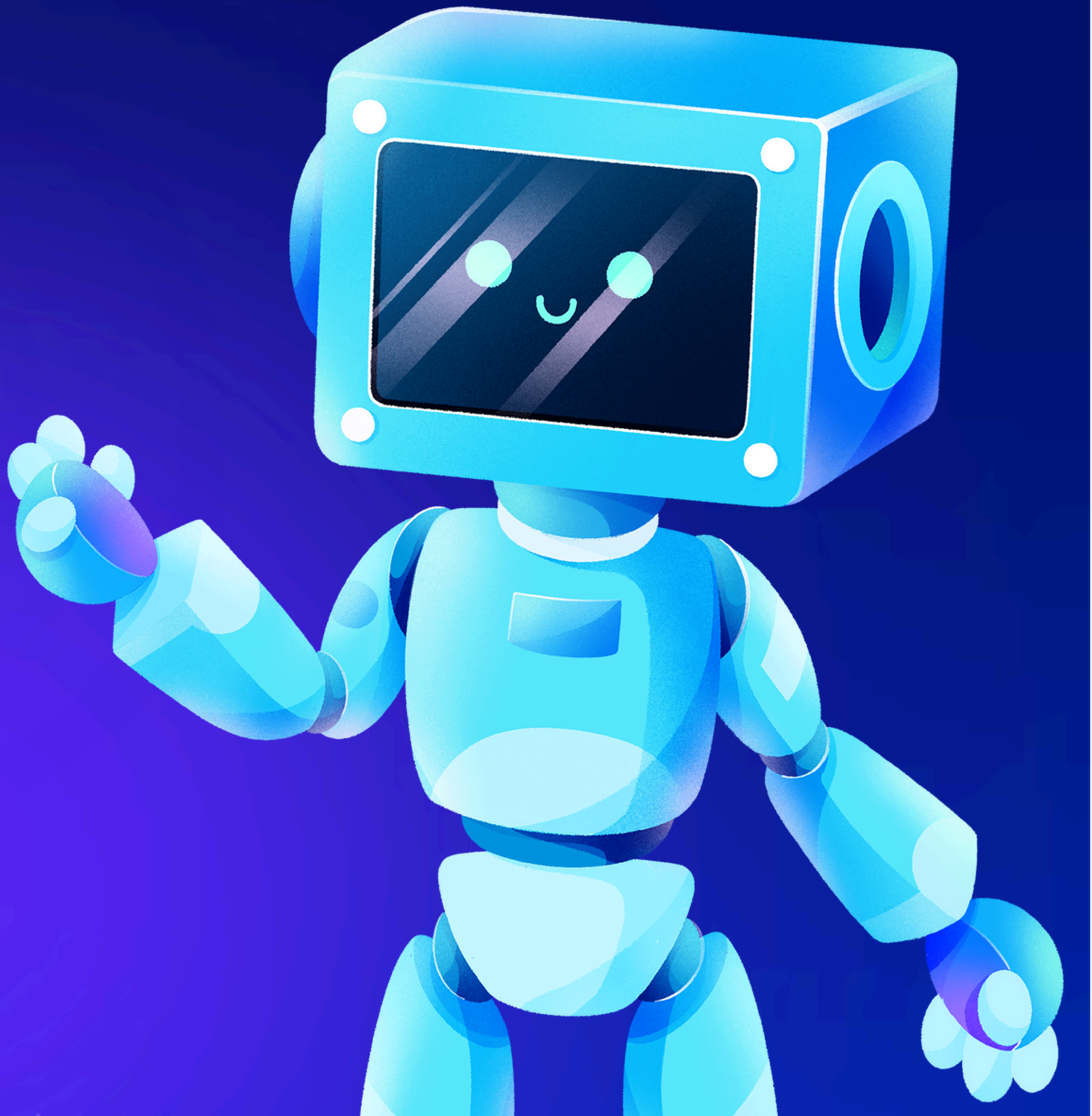


MACHINE LEARNING

PROJECT

BY GROUP 6 :

MONICA - KHUSHI - GREG - NAPAT



INTRODUCTION

Forecasting Turnover in the Australian Retail Sector is essential for optimising stock levels and responding proactively to market changes.

Through leveraging machine learning, our project will provide a reliable solution and helps business make informed decision.

OBJECTIVES

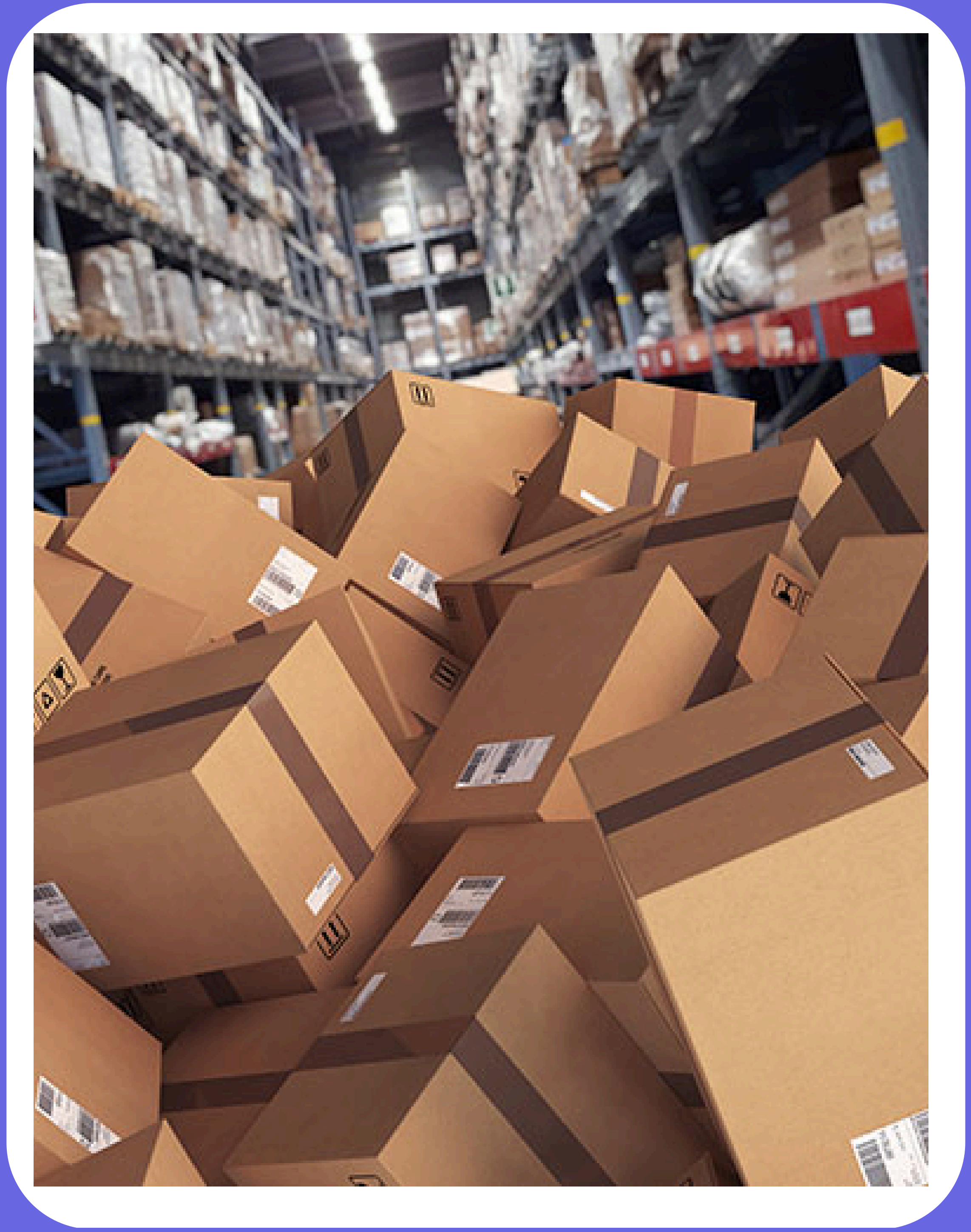
- To analyze historical turnover data to identify trends in the Australian retail sector.
- To visualize and interpret the results for actionable insights that can inform retail strategies.
- To develop a machine learning model and use that to forecast turnover across the retail categories for the dataset.

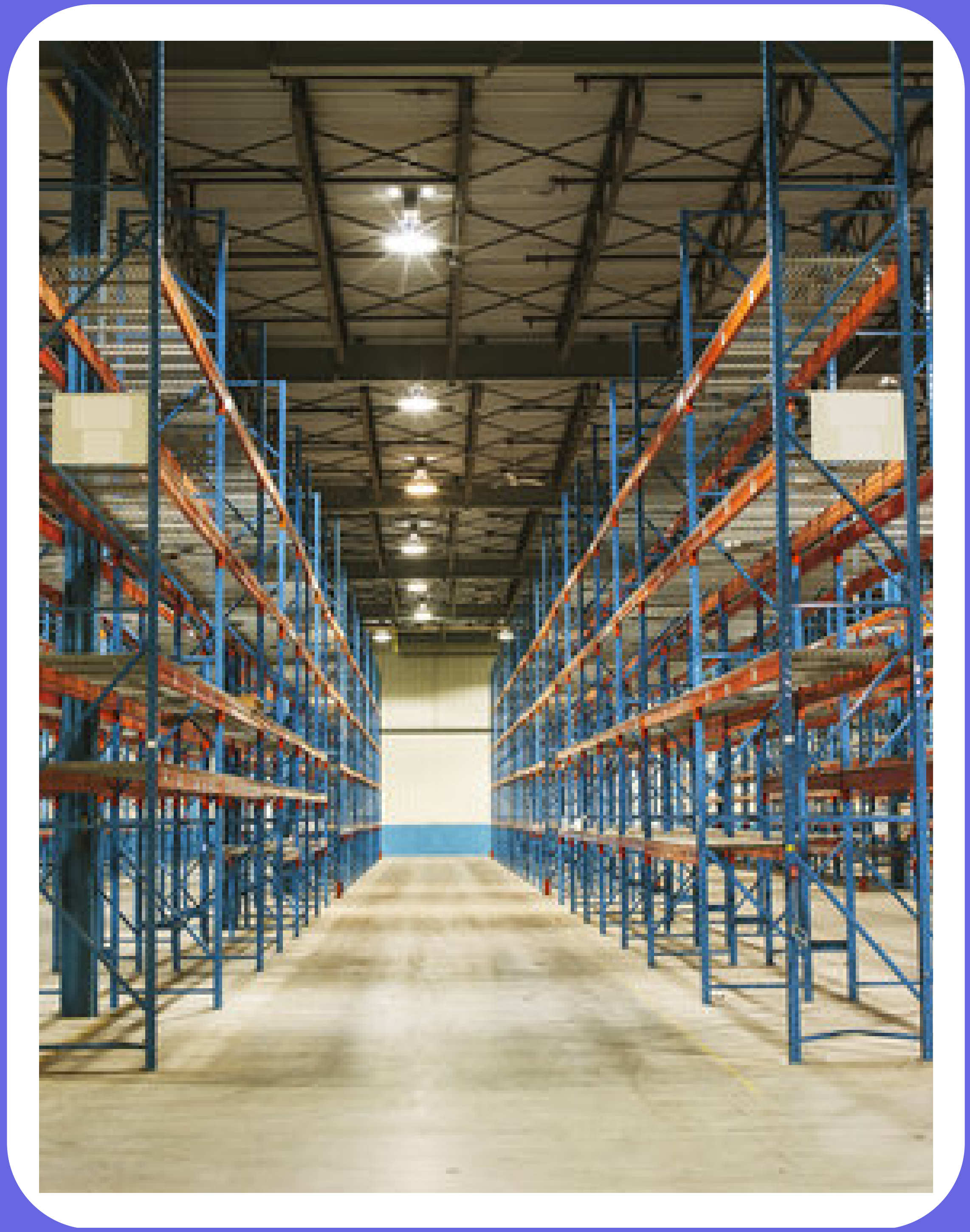
INEFFICIENCIES OF STOCK ORDERING

KEY CHALLENGES:

- Overstocking: Leads to increased inventory holding costs, potential product expiration and waste.

- Understocking: Results in missed sales opportunities, loss of customer trust and reputational damage.

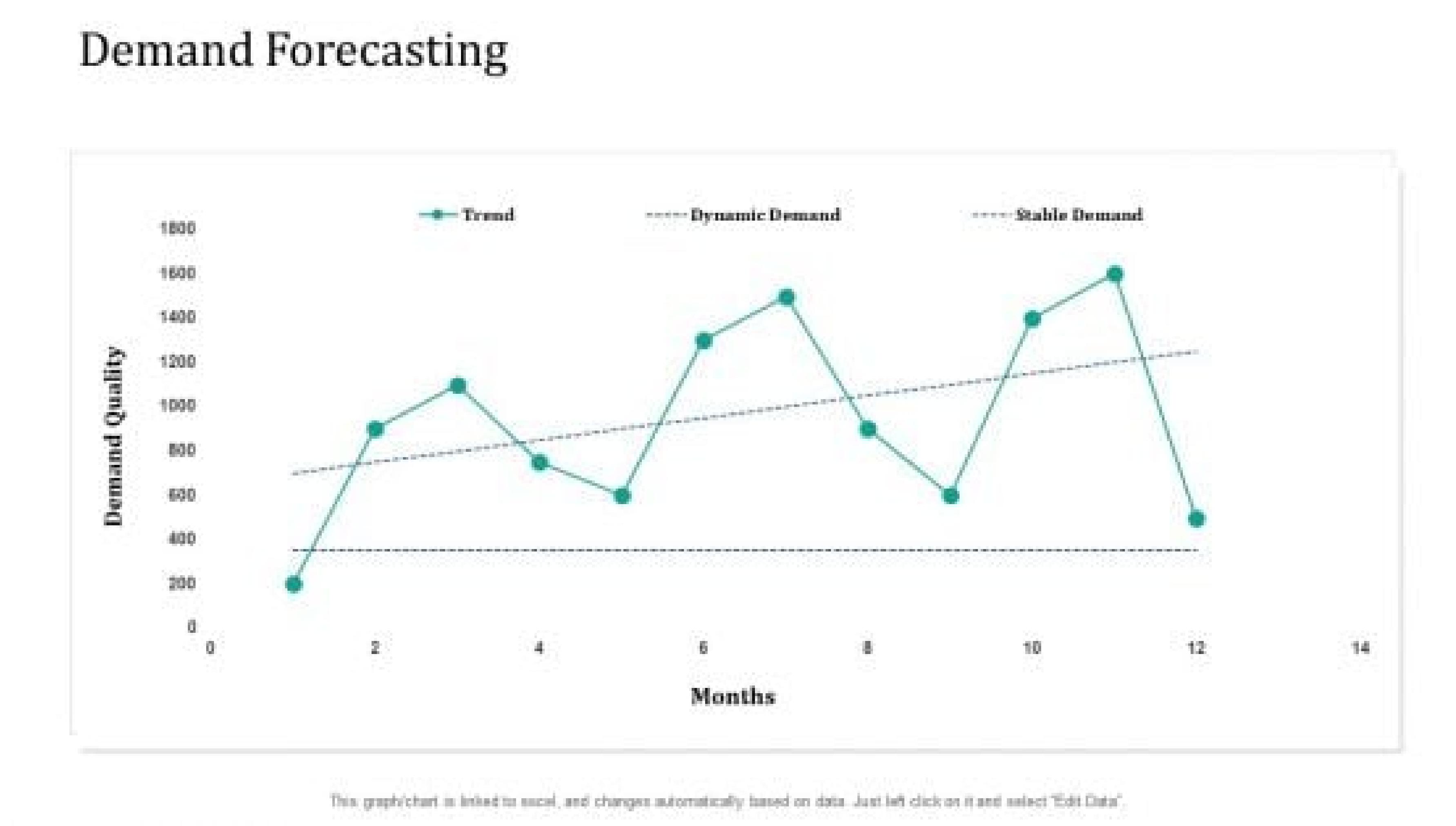




IMPACT OF ACCURATE DEMAND FORECASTING

- Optimised Resource Allocation: Enables better planning for logistics, staffing and warehousing

- Reveals Seasonal Trends
- Helps the business rationalise its cash flow
- Assist the business in planning the supply chain
- Reveals how external factors may influence your sales
- Prepares the business for the future



DATA OVERVIEW

Source: www.abs.gov.au

- The dataset contains various columns related to retailing across different categories, such as:

- Original : Raw data values.
- Seasonally : Adjusted for seasonal effects.
- Trend : Long-term trends in the data.
- Total : Total turnover for each of the above aspects.

(509, 22)

	year	Original_Food retailing	Original_ Household goods retailing	Original_Clothing, footwear and personal accessory retailing	Original_Department stores	Original_ Other retailing	Original_Cafes, restaurants and takeaway food services	Original_Total (Industry)
0	Apr-1982	1162.6	592.3	359.9	460.1	479.1	342.4	3396.4
1	May-1982	1150.9	629.6	386.6	502.6	486.1	342.1	3497.9
2	Jun-1982	1160.0	607.4	350.5	443.8	467.5	328.7	3357.8
3	Jul-1982	1206.4	632.4	359.3	459.1	491.1	338.5	3486.8
4	Aug-1982	1152.5	622.6	325.2	438.4	485.7	331.5	3355.9

```
year                                                             0
Original_Food retailing                                          0
Original_ Household goods retailing                              0
Original_Clothing, footwear and personal accessory retailing     0
Original_Department stores                                       0
Original_ Other retailing                                        0
Original_Cafes, restaurants and takeaway food services           0
Original_Total (Industry)                                        0
Seasonally_Food retailing                                        0
Seasonally_Household goods retailing                             0
Seasonally_Clothing, footwear and personal accessory retailing   0
Seasonally_Department stores                                     0
Seasonally_Other retailing                                       0
Seasonally_Cafes, restaurants and takeaway food services         0
Seasonally_Total (Industry)                                      0
Turnover ;  Total (State) ;  Food retailing ;                   28
Trend_Household goods retailing                                 28
Trend_Clothing, footwear and personal accessory retailing       28
Trend_Department stores                                         28
Trend_Other retailing                                           28
Trend_Cafes, restaurants and takeaway food services             28
Trend_Total (Industry)                                          28
dtype: int64
```

METHODOLOGY

01 Data Collection

02 Data Preparation

03 Data Cleaning

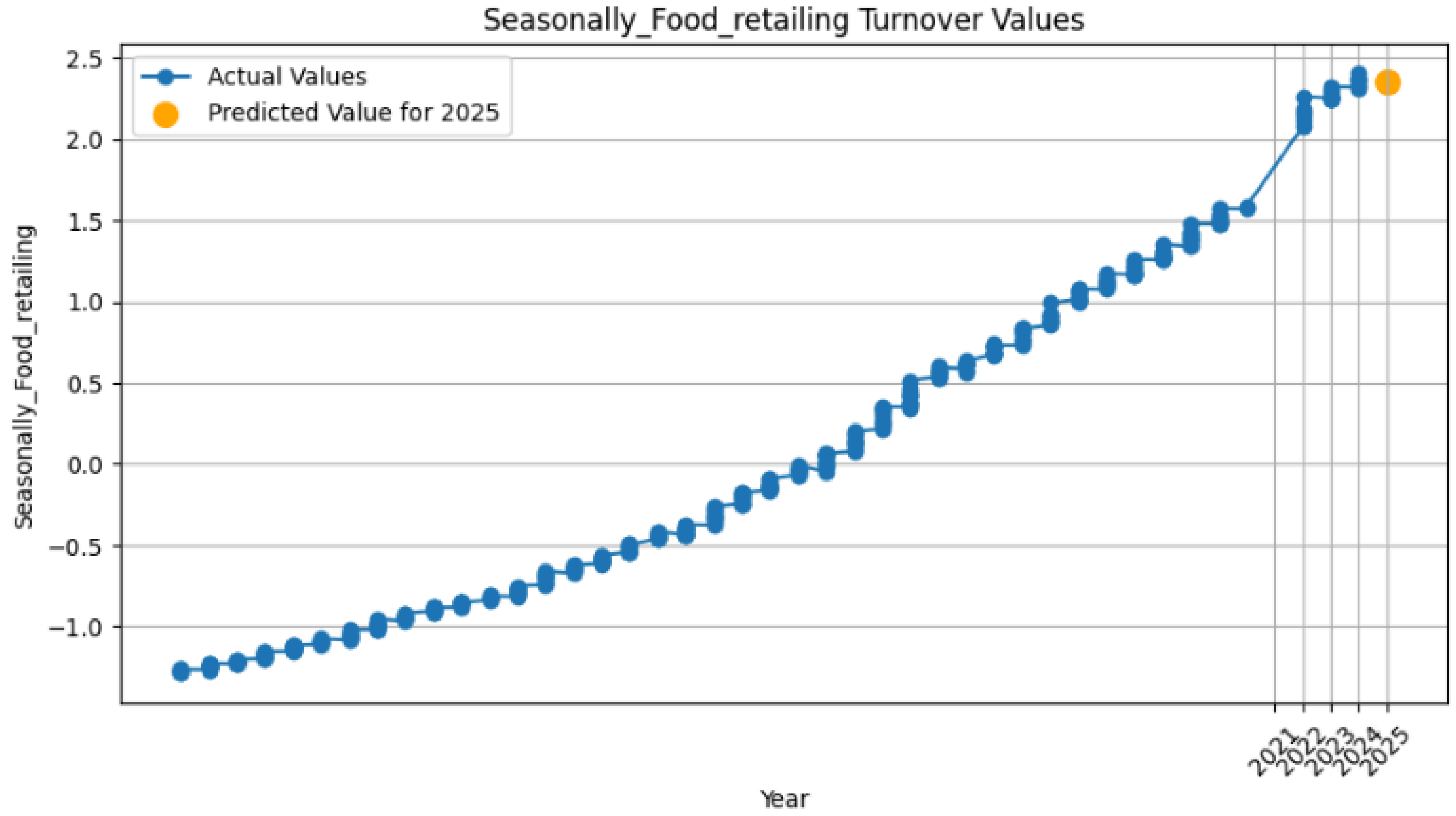
04 Split and Transform the Data for Machine Learning

05 Model Trials

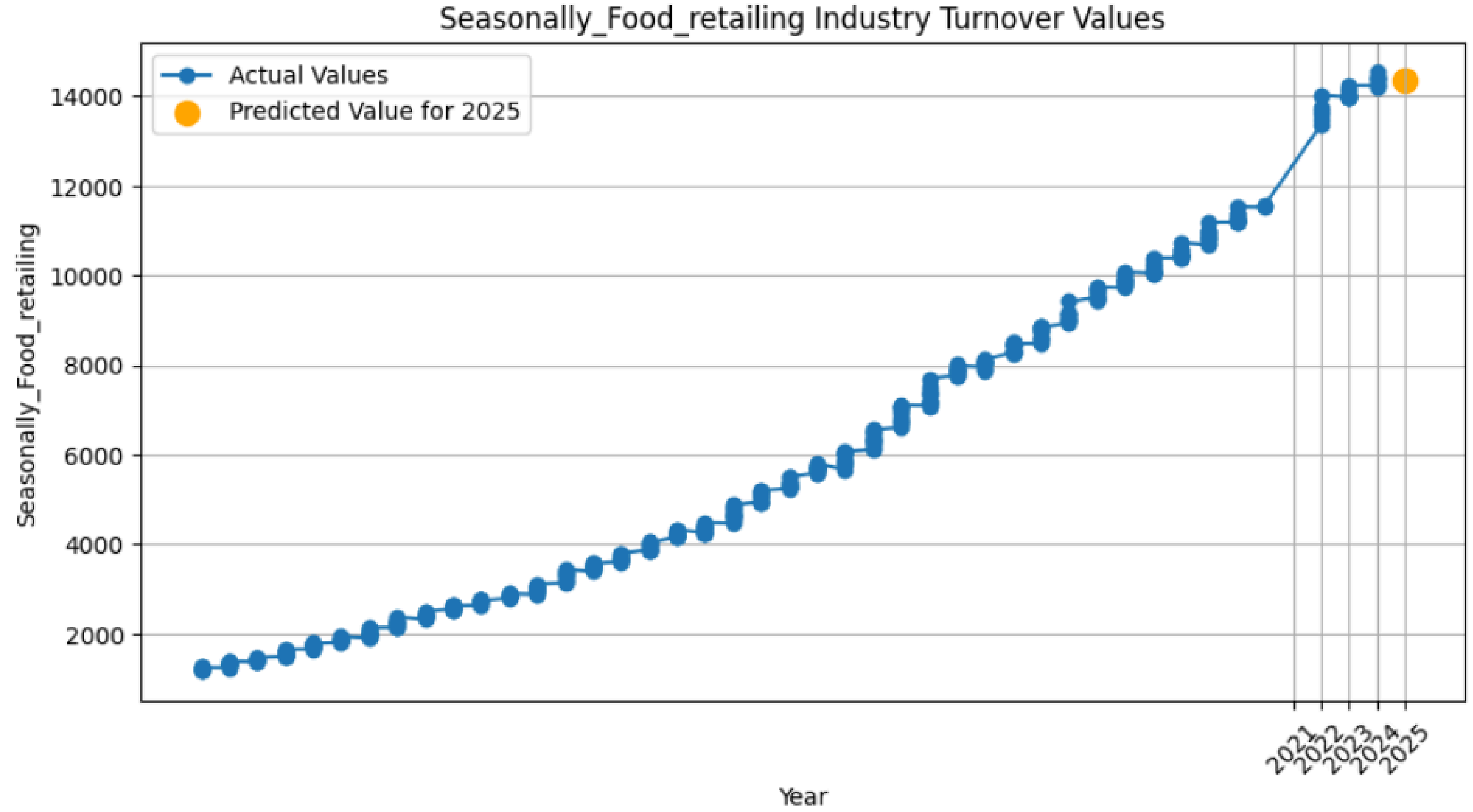
06 Final Model Selection

RANDOM FOREST REGRESSOR MODEL EVALUATION

```
Evaluation Metrics:
MAE: 0.024583740560147505
MSE: 0.0010722544279414376
R²: 0.9991031918244959
RMSE: 0.032745296272005814
Predicted Turnover for Seasonally_Food_retailing for 2025: 2.35
```



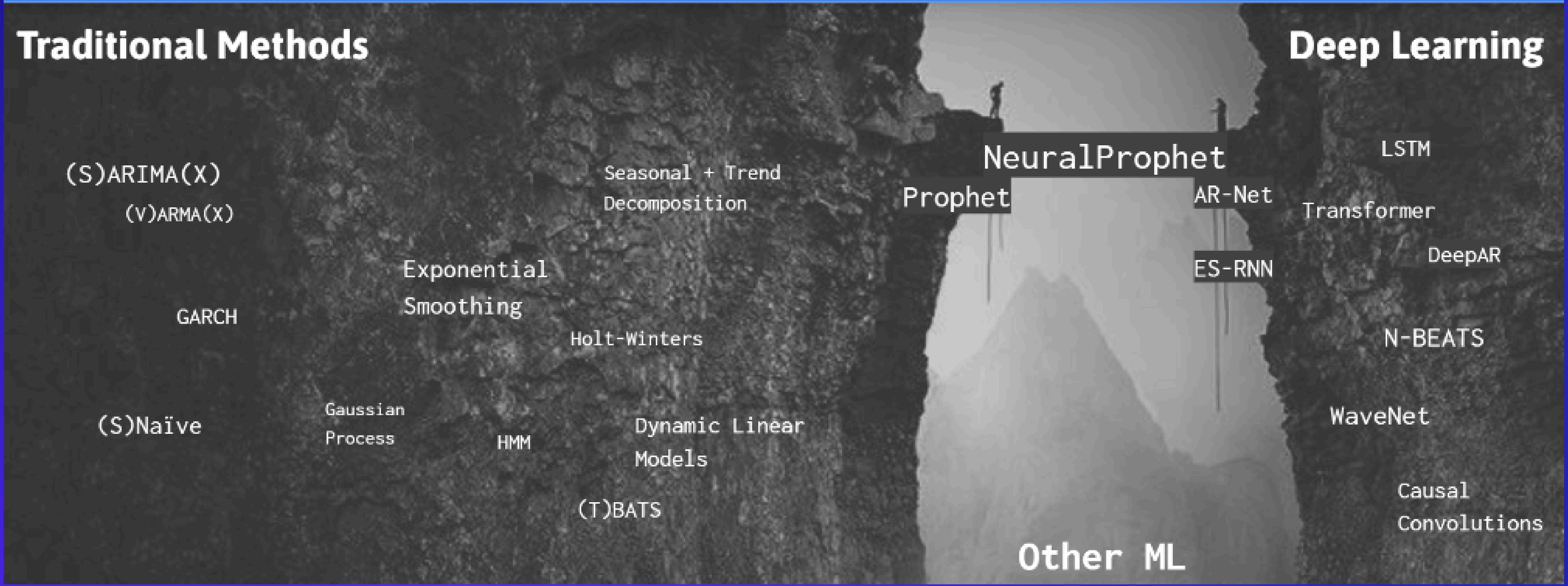

```
Evaluation Metrics:
MAE: 89.09938520604982
MSE: 14084.821098089518
R²: 0.9991031918244959
RMSE: 118.67948895276521
Predicted Turnover for Seasonally_Food_retailing for 2025: 14360.65
```



Random Forest Model Evaluation Matrics for Seasonally Adjusted Data

	Food Retailing	Household Goods	Clothing & Accessories	Department Stores	Other Retailing	Cafes & Restaurants	Total Industry
MAE	89.099385	54.459772	26.437811	26.561176	35.220297	37.054921	209.669225
MSE	14084.821098	6521.164924	1534.238286	1797.280127	2081.140878	2711.381365	78690.374617
R2	0.999103	0.997387	0.997189	0.989673	0.998989	0.998748	0.999138
RMSE	118.679489	80.75373	39.169354	42.394341	45.619523	52.070926	280.518047
2024 Turnover	14237.2	5753.8	2948.4	1878.1	5529.8	5379.0	35726.1
2025 Prediction	14360.65	5769.68	3013.27	1896.16	5621.78	5396.39	36057.90

Time series forecasting is messy.
We need hybrid models to bridge the gap.



Source: NeuralProphet GitHub
https://github.com/ourownstory/neural_prophet/blob/main/notes/NeuralProphet_Introduction.pdf

NEURAL PROPHET MODEL

Neural Prophet is a time-series forecasting model, based on PyTorch and inspired by Facebook Prophet and AR-Net.
It combines Neural Networks and traditional time-series algorithms.

INITIAL CONFIGURATION

- Input features - Cleaned Turnover data
- Data frequency 'Month Start'

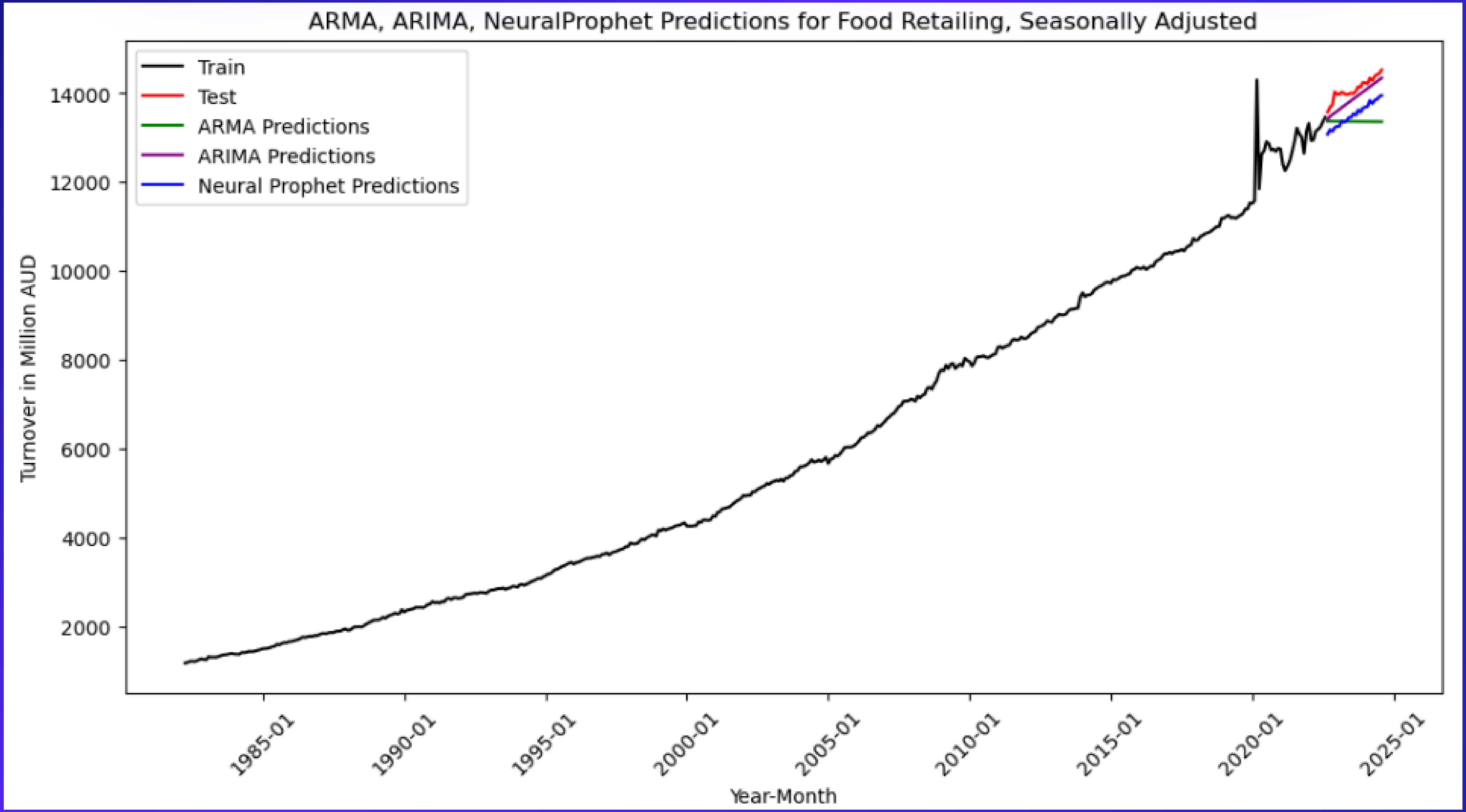
```python
# Create and train NeuralProphet model
np_model = NeuralProphet()
metrics = np_model.fit(np_train, freq='MS', epochs=300)
```

MODEL TRAINING

- Train on historical data (up to forecast date)
- Forecast (test) last 2 years of turnover
- Epochs - was initially set to automatic (150) and later changed to higher values to evaluate for improved prediction performance.

```
Epoch 300: 100% 300/300 [00:00<00:00, 99737.73it/s]
Finding best initial lr: 100% 219/219 [00:54<00:00, 55.22it/s]
Predicting DataLoader 0: 100% 1/1 [00:00<00:00, 221.18it/s]
```

MODEL PREDICTIONS COMPARISON (1)



MODEL PREDICTIONS COMPARISON (2)

Error statistics when predicting the last 24 months of data (prediction vs actual)

ARMA

Category	Food	Household_Goods	Clothing_and_Footware	Department_Stores	Other_Retailing	Cafes_Restaurants_TakeAway
RMSE	772.989309	292.595682	125.613863	46.045513	138.062098	338.316832
R-squared	-9.889264	-3.791624	-6.575085	-0.524058	-0.386605	-11.686900

ARIMA

Category	Food	Household_Goods	Clothing_and_Footware	Department_Stores	Other_Retailing	Cafes_Restaurants_TakeAway
RMSE	187.362941	511.476138	71.555975	53.478189	359.990876	164.790027
R-squared	0.360237	-13.641907	-1.458126	-1.055798	-8.427291	-2.010025

Neural Prophet

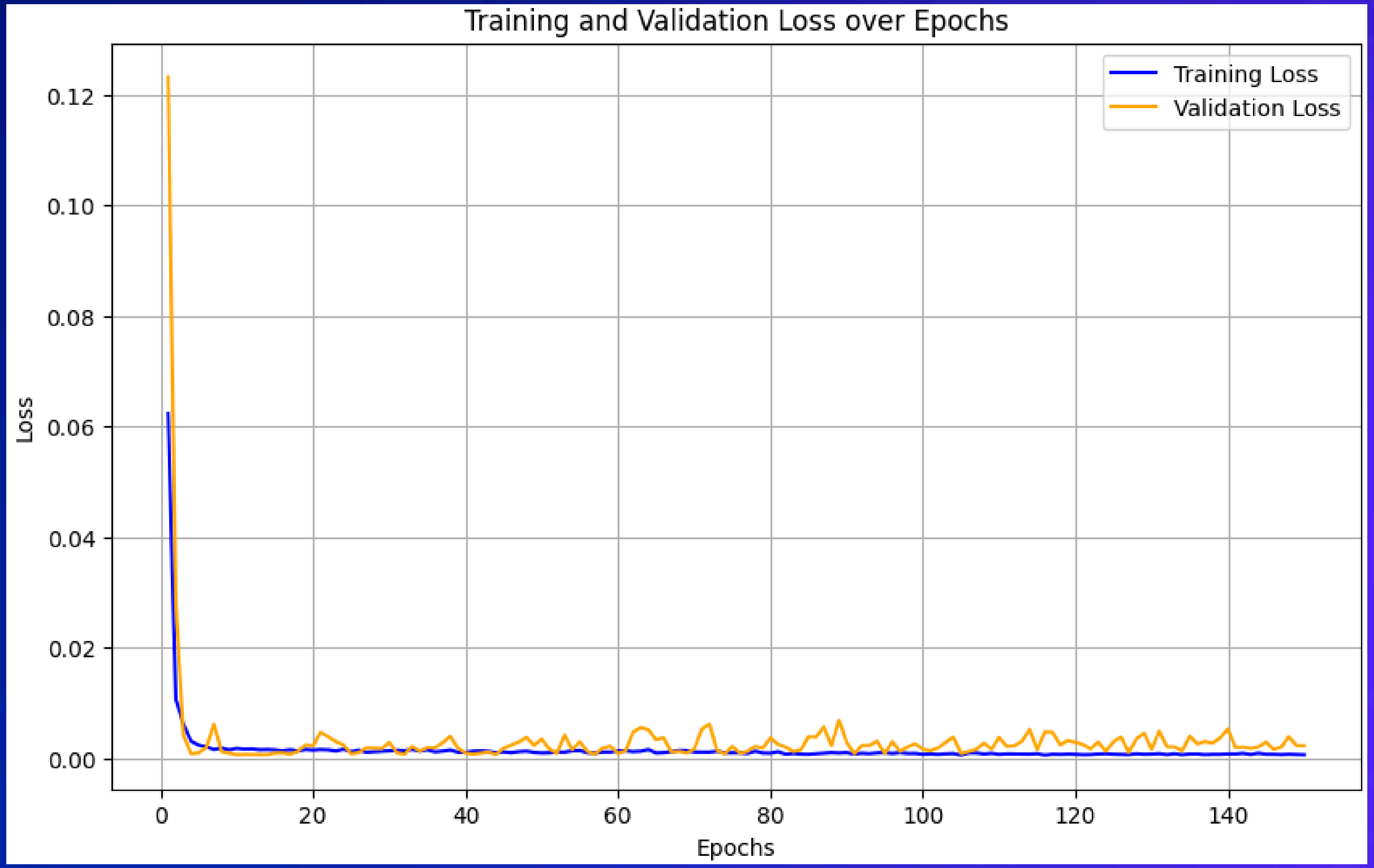
Category	Food	Household_Goods	Clothing_and_Footware	Department_Stores	Other_Retailing	Cafes_Restaurants_TakeAway
RMSE	602.736023	236.609438	433.562349	244.886751	493.045695	998.397975
R-squared	-5.620730	-2.133365	-89.243405	-42.107966	-16.683910	-109.488142

UNIVARIATE LSTM MODEL

LSTM is a type of recurrent neural networks (RNN)
Specifically designed to handle sequence of data
making it suitable for time series forecasting

```python
# Build the LSTM model with hyperparameter tuning and dropout

model = keras.Sequential([
    layers.LSTM(100, activation='relu', return_sequences=True, input_shape=(X_train.shape[1], 1)),
    layers.Dropout(0.2),
    layers.LSTM(50, activation='relu'),
    layers.Dropout(0.2),
    layers.Dense(1)
])
```



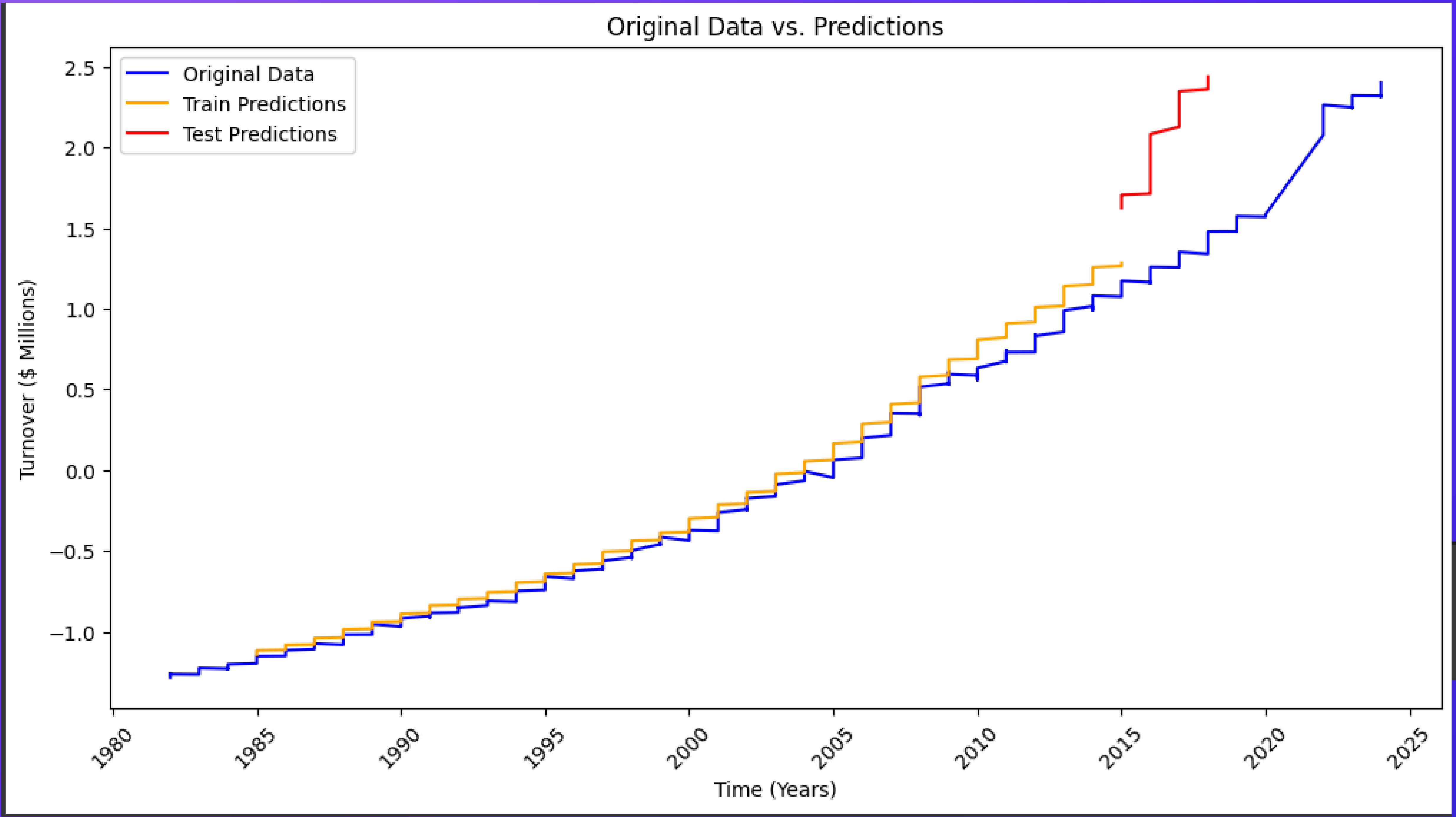
INITIAL CONFIGURATION

- Input features - Cleaned Turnover data
- Model architecture - Single layer LSTM with dropout

MODEL TRAINING

- Layer structure is composed of 100 LSTM units to capture underlying patterns in the data
- Included dropout layer to prevent overfitting
- Adam optimizer was chosen
- Epochs - was initially set to 100 and later changed to 150 to increase performance

MODEL PERFORMANCE



Performance Metrics

- Mean Absolute Error (MAE)
- Mean Squared Error
- R Squared Value

```
Mean Absolute Error: 0.1339917084728732
Mean Squared Error: 0.0241582276775032
R-squared: 0.8195625273484578
```

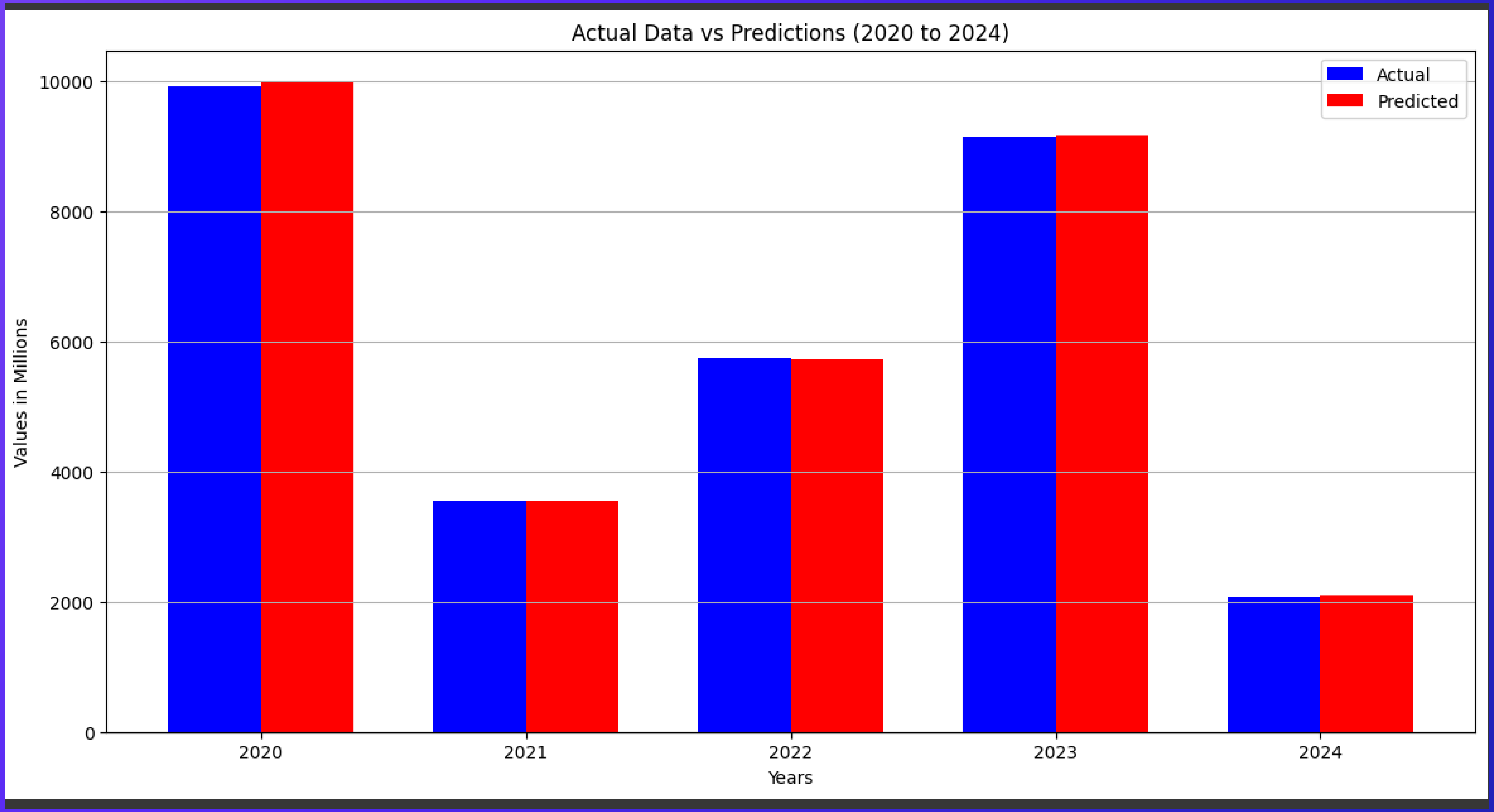
ENHANCED LSTM MODEL PERFORMANCE

Advanced Architectures

1) Random Split between train-test was used

2)Bidirectional LSTM
This allows the Model to learn from past and future

3) Stacked LSTM
Multiple LSTM layers stacked on top of each other help the model learn more complex features.




```
Mean Squared Error: 0.0001851321382156742
Mean Absolute Error: 0.009500049060792744
R-squared: 0.9998433564198007
```

RESULTS & RECOMMENDATIONS

The preferred model was LSTM

- It produced better results compared to other Models
- It works well with sequential data, making it ideal for time series forecasting.
- LSTMs have memory cells that can maintain information over long period

LSTM Model Evaluation Matrics for Seasonally Adjusted Data

	Food Retailing	Household Goods	Clothing & Accessories	Department Stores	Other Retailing	Cafes & Restaurants	Total Industry
MAE	35.33	43.93	27.75	30.62	22.17	22.11	131.10
MSE	2319.54	7897.00	2641.52	3588.07	1190.67	997.98	50956.20
R2	1.00	1.00	1.00	0.98	1.00	1.00	1.00
RMSE	48.16	88.87	51.40	59.90	34.51	31.59	225.73
2024 Turnover	14237.20	5753.80	2948.40	1878.10	5529.80	5379.00	35726.10
2025 Prediction	14595.79	5846.76	3075.35	1936.21	5775.18	5460.77	36727.02

- Recommendations for further work include:
  - Investigate model parameter optimisation with tools such as Optuna (https://optuna.org/)
  - Explore models with attention mechanisms (like Transformers) it can capture long-range dependencies in time series data more effectively

THANK YOU!