AI IN SALES PREDICTION AND FORCAST ANALYSIS

Contributors:

1. Spandan Dhadse
2. Akshay Pankar
3. Metturu Mouli
4. Advait Rahul Ghatge
5. Bhodigam Akshitha

Date: 24/05/23

1. ABSTRACT

All industries aim to manufacture just the right number of products at the right time, but for retailers this issue is particularly critical as they also need to manage perishable inventory efficiently. Too many items and too few items are both scenarios that are bad for business. Most of the business organizations heavily depend on a knowledge base and demand prediction of sales trends. The accuracy in sales forecast provides a big impact in business. Companies of all sizes can now harness machine learning to anticipate sales trends without the need for data scientists.

In this model, all inputs are pre-processed to be understandable by the machine. This is a linear regression model based on supervised learning, so the output will be provided along with the input. Then inputs are then fed to the model along with desired output. The model will plot(learn) a relation(function) between the input and output. This function or relation is then used to predict the output for a specific set of inputs to an mobile application. Input parameters like date and previous sales are labelled as input, and the number of sales is marked as output.

2. INTRODUCTION:

Sales forecasting, a vital aspect of business strategy, allows companies to stay ahead, make informed choices, and predict upcoming trends. Traditional forecasting methods are often time-consuming, prone to errors, and subject to human inaccuracies. With the emergence of machine learning and no-code predictive analytics, how businesses forecast sales has been transformed.

Businesses can quickly generate precise forecasts, monitor trends, and make well-informed decisions. By using machine learning algorithms to identify patterns and trends, businesses can anticipate shifts in the market and stay ahead of the competition.

3. PROBLEM STATEMENT

- Predicting sales of a company needs time series data of that company and based on that data the model can predict the future sales of that company or product.

- Looking at the various Stores Sales around the world are tasked with predicting their daily sales in advance.

- Store sales are influenced by many factors, including promotions, competition, school and state holidays, seasonality, and locality. With thousands of sales records we will be predicting sales based on their unique circumstances, the accuracy of results can be quite varied.

4. MARKET/CUSTOMER/BUSSINESS NEED ASSESSMENT

Demand forecasting is the process of predicting what the demand for certain products will be in the future. This helps manufacturers to decide what they should produce and guides retailers toward what they should stock.

Demand forecasting is aimed at improving the following processes:

- Supplier relationship management
- Customer relationship management
- Order fulfilment and logistics

- Marketing campaigns
- Manufacturing flow management
- Using Machine Learning to Predict Customers' Next Purchase Day

5. TARGET SPEIFICATION AND CHARACTERIZATION

A. Our main objective is industry and big marts even small-scale shops and marts, where for publicly traded companies, accurate forecasts confer credibility in their market.

B. Companies of all sizes can now harness machine learning to anticipate sales trends without the need for data scientists.

C. Traditional forecasting methods are often time-consuming, prone to errors, and subject to human inaccuracies.

D. For retailers this issue is particularly critical as they also need to manage perishable inventory efficiently. Too many items and too few items are both scenarios that are bad for business. So, accurate sales forecast needs to be achieved.

E. What do you hope to achieve in the next month? Year? 5-years?

F. How many customers do you hope to have next month and next year?

G. How much will each customer hopefully spend with your company?

6. EXTERNAL SEARCH (INFORMATION SOURCES)

Data Set Used:
The dataset used is provided in the below link.
Data Exploring:

```
train.head()
```

	date	store	item	sales
0	2013-01-01	1	1	13
1	2013-01-02	1	1	11
2	2013-01-03	1	1	14
3	2013-01-04	1	1	13
4	2013-01-05	1	1	10

```
test.head()
```

	id	date	store	item
0	0	2018-01-01	1	1
1	1	2018-01-02	1	1
2	2	2018-01-03	1	1
3	3	2018-01-04	1	1
4	4	2018-01-05	1	1

Data Cleaning:

Cleaning the data

null value analysis

```
train.isnull().sum()
```

```
date     0
store    0
item     0
sales    0
dtype: int64
```

```
test.isnull().sum()
```

```
id       0
date     0
store    0
item     0
dtype: int64
```

Data pre-processing:

Data pre-processing refers to the technique of preparing (cleaning and organizing) the raw data to make it suitable for a building and training Machine Learning models.

```
train.head()
```

	date	store	item	sales	year	month	day	weekday
0	2013-01-01	1	1	13	2013	1	1	1
1	2013-01-02	1	1	11	2013	1	2	2
2	2013-01-03	1	1	14	2013	1	3	3
3	2013-01-04	1	1	13	2013	1	4	4
4	2013-01-05	1	1	10	2013	1	5	5

```
test.head()
```

	id	date	store	item	year	month	day	weekday
0	0	2018-01-01	1	1	2018	1	1	7
1	1	2018-01-02	1	1	2018	1	2	1
2	2	2018-01-03	1	1	2018	1	3	2
3	3	2018-01-04	1	1	2018	1	4	3
4	4	2018-01-05	1	1	2018	1	5	4

Market Segmentation and Analysis:

Why is sales forecasting important?

Sales forecasting is a critical process for companies, providing valuable insights into future sales performance. It holds significant importance for both privately held and publicly traded companies. Accurate forecasts instill confidence in privately held companies and confer credibility in the market for publicly traded companies.

Sales forecasting has a wide-ranging impact across different departments within an organization. Finance departments rely on forecasts to develop budgets and make informed decisions regarding capacity plans and hiring. Production teams use sales forecasts to plan their production cycles effectively. Sales operations benefit from forecasts for territory and quota planning, while supply chain departments utilize them for material purchases and production capacity planning.

However, many companies face challenges due to disconnected methodologies and poor information sharing. This lack of communication can lead to adverse business outcomes. For instance, if product marketing creates demand plans that are misaligned with sales quotas or attainment levels, it can result in inventory imbalances and inaccurate sales targets, negatively impacting the company's profitability.

To avoid such costly mistakes, companies must prioritize regular and high-quality sales forecasting. This commitment involves fostering effective communication and collaboration among departments. By sharing information and aligning strategies, companies can optimize inventory levels, set accurate sales targets, and make informed decisions that drive growth and profitability.

Benefits of having an efficient sales forecast:

An accurate sales forecast process offers many benefits. These include:

- **Enhanced decision-making:** Accurate sales forecasts provide valuable insights for making informed decisions about the future of the business.

- **Mitigation of sales pipeline and forecast risks:** By accurately predicting sales, companies can reduce the uncertainties and risks associated with their sales pipeline and forecasting process.

- **Time-saving in planning and assignments:** With accurate sales forecasts, less time is spent on planning territory coverage and setting quota assignments since the forecasts provide a reliable foundation for these tasks.

- **Benchmarking for future trend assessment: Accurate** forecasts serve as benchmarks, allowing companies to assess trends and patterns in their sales performance over time, facilitating better strategic planning.

- **Focus on high-revenue opportunities**: Accurate sales forecasts enable companies to identify and prioritize high-revenue and high-profit sales opportunities, leading to improved win rates and overall sales performance.

- **Avoiding loss:** Sales prediction and forecast enhances the future planning by accounting all the factors which needs to be improved or evaluated in order to avoid the loss.

7. BENCHMARKING

Benchmarking sales forecasts means comparing them with other relevant data points, such as historical trends, industry averages, market share, customer feedback, or competitor performance. Benchmarking can help to validate your assumptions, identify gaps and opportunities, and adjust your strategies and tactics accordingly.

- For example, if your sales forecasts are much higher than your industry average, you might want to investigate the reasons behind your optimism and test your hypotheses. On the other hand, if your sales forecasts are much lower than your competitors, you might want to explore the factors that are holding you back and find ways to overcome them.

8. CONCEPT GENERATION

- For analysis of sales and their prediction and forecast is done by evaluating five popular Machine Learning models and obtained different accuracy rate depending up on the data provided.

The five Machine Learning models used are:

1. Seasonal Naive Model

2. Holt Winter's Triple Exponential Smoothing Model

3. ARIMA Model: Autoregressive Integrated Moving Average

4. Supervised Machine Learning: Linear Regression Model

5. Random Forest Model

1.Seasonal Naive Model

The seasonal naive method is a simple forecasting technique that assumes the future values will be the same as the last observed value from the same season. In the case of sales forecasting, it assumes that sales in a future period will be equal to the sales in the same period of the previous year.

Methodology:

Seasonal Decomposition: The first step in the seasonal naive model is to decompose the historical sales data into its different components: trend, seasonality, and random variation

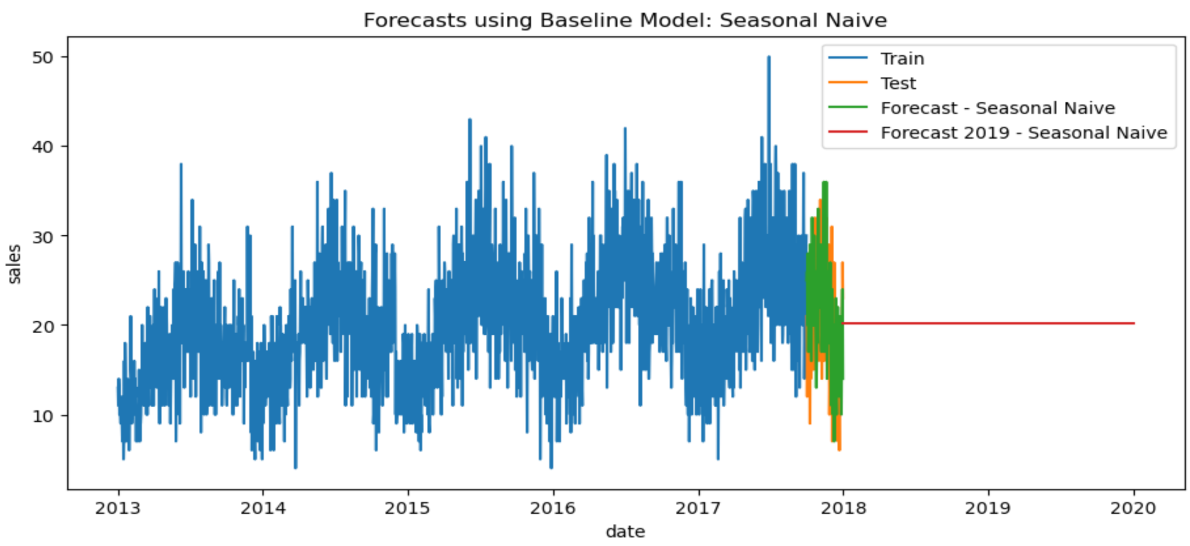
Seasonal Adjustment: Once the seasonal component has been isolated, it is adjusted to remove its influence from the data.

Forecasting: With the deseasonalized data, the seasonal naive model employs a simple approach of using the value from the previous season (i.e., the same season of the previous year) as the forecast for the corresponding season in the future.

Seasonal Adjustment and Re-seasonalization: After obtaining the forecasted values for each season, the seasonal component that was initially removed is added back to the forecasts to reintroduce the seasonal pattern.

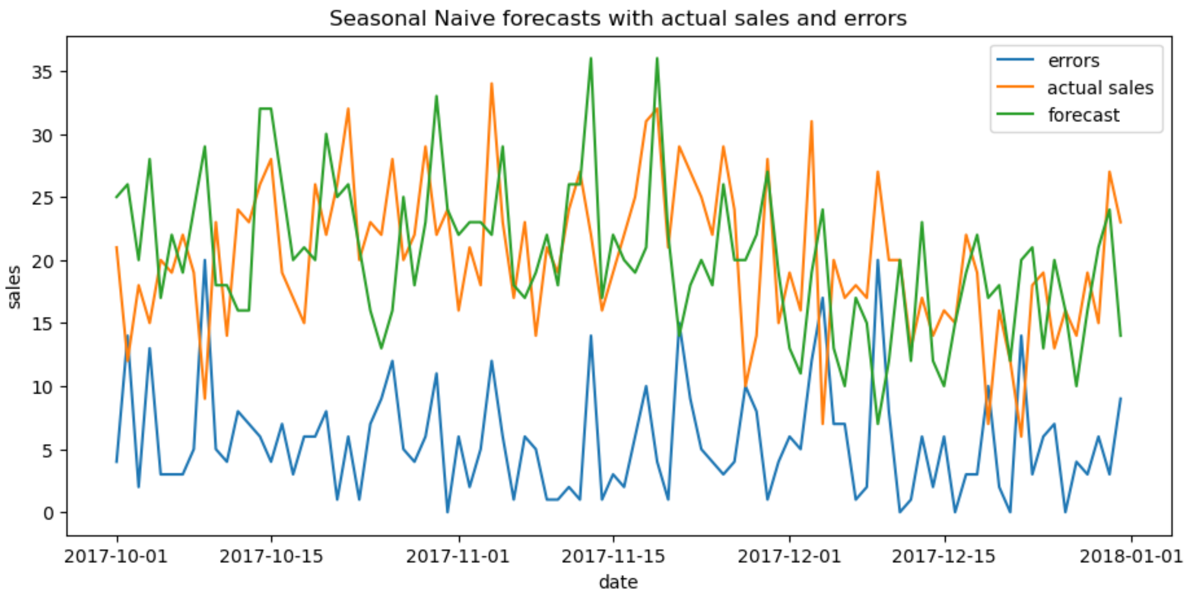
Final Sales Forecast: The seasonal naive model provides the final sales forecast by combining the seasonally adjusted forecasts with the seasonal component.

Prediction and Forecast of Sales:



➢ Analysis of prediction and forecast is done on the test data and forecast is done up to next two years.

Actual Sales and Errors:



Error Data Frame:

model	total_sales	total_sn_pred_sales	overall_error	MAE	RMSE	MAPE
Seasonal Naive	1861	1857	4	5.630435	7.130766	27.834498

2. Holt Winter's Triple Exponential Smoothing Model

Holt-Winters' Triple Exponential Smoothing (TES) is a widely used forecasting model for sales prediction and forecasting. It is particularly effective when dealing with time series data that exhibit trends, seasonality, and irregular fluctuations. The model extends the simple exponential smoothing method by incorporating three components: level, trend, and seasonality.

Methodology:

Level: The level component represents the average value of the sales data. It is updated at each time step by combining the current observation with the level estimate from the previous time step.

Trend: The trend component captures the direction and rate of change in the sales data. It is updated based on the current observation and the trend estimate from the previous time step.

Seasonality: The seasonality component accounts for recurring patterns or seasonal fluctuations in the sales data. It involves identifying the length of the seasonal pattern (e.g., monthly, quarterly).

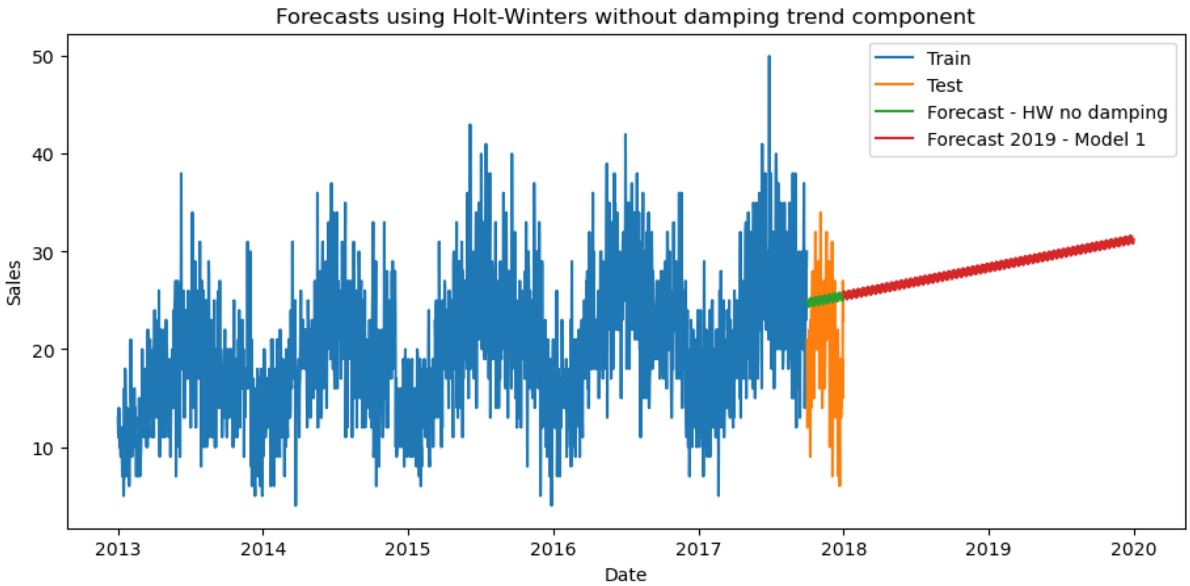
Smoothing: The TES model employs smoothing factors (alpha, beta, and gamma) to determine the weights given to the level, trend, and seasonality components, respectively.

Forecasting: The TES model utilizes the level, trend, and seasonality components to generate forecasts. The seasonality component is used to adjust the forecasted values to account for seasonal variations.

Updating: As new sales data becomes available, the TES model updates the level, trend, and seasonality components to adapt to any changes in the underlying patterns.

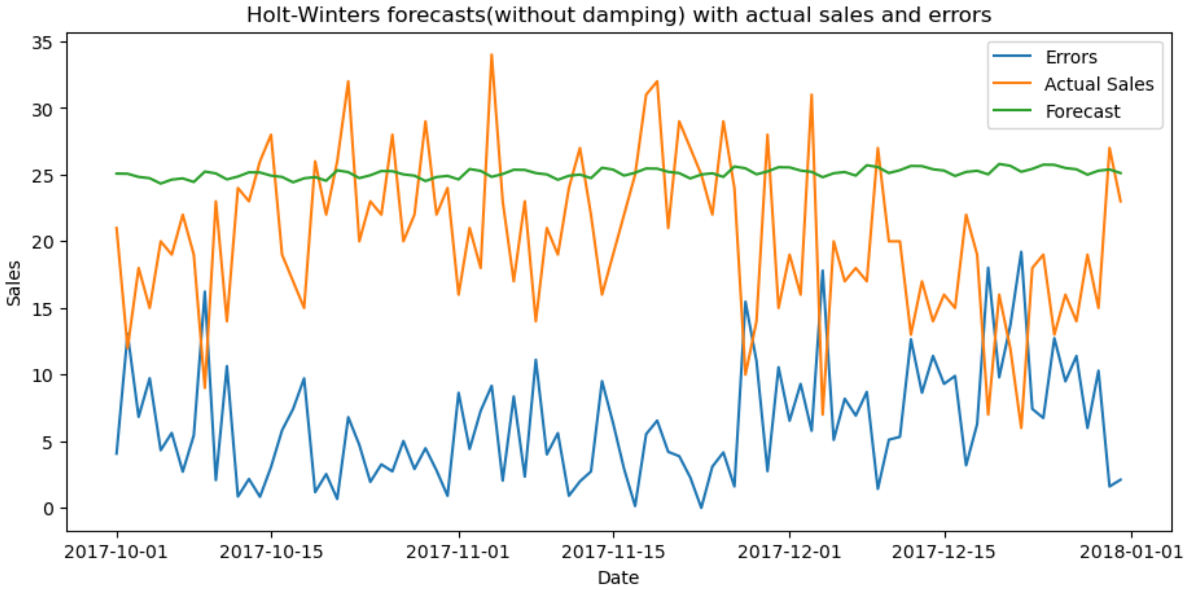
Prediction and Forecast of Sales:

Forecasts using Holt-Winters without damping trend component:



- ➢ Analysis of prediction and forecast is done on the test data and forecast is done up to next two years.

Actual Sales and Errors:

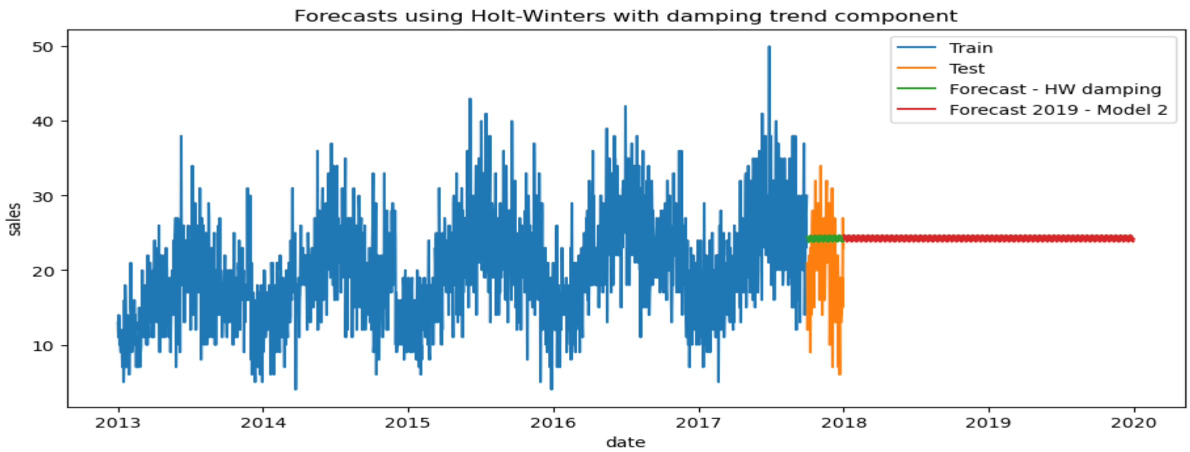


Error Data Frame:

model	total_sales	total_pred_sales	holt_winters_overall_error
Holt-Winters	1861	2310.834517	- 449.834517

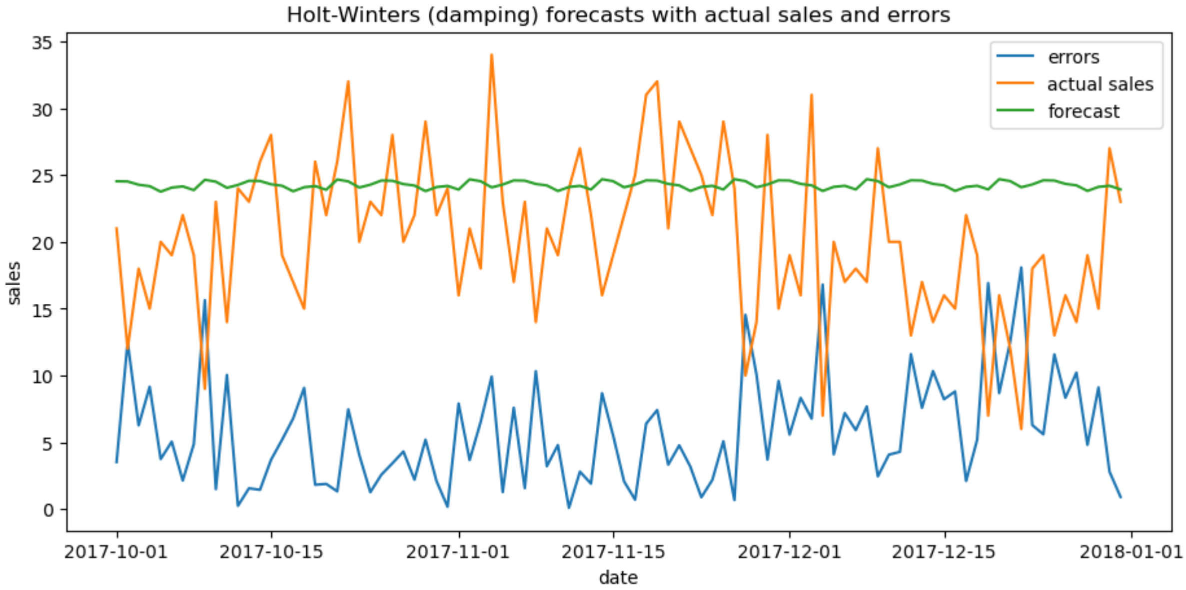
MAE	RMSE	MAPE
6.302484	7.644371	31.156824

Forecasts using Holt-Winters with damping trend component:



- Analysis of prediction and forecast is done on the test data and forecast is done up to next two years.

Actual Sales and Errors:



Error Data Frame:

model	total_sales	total_pred_sales	holt_winters_overall_error
Holt-Winters	1861	2232.198699	-371.198699

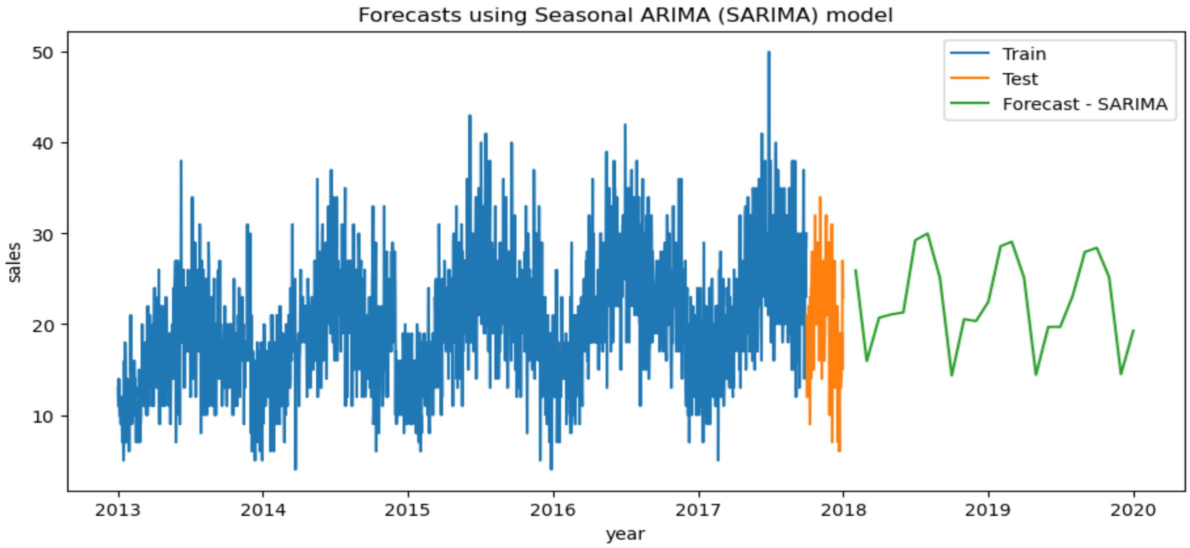
MAE	RMSE	MAPE
5.804361	7.08568	28.694314

3. ARIMA Model: Autoregressive Integrated Moving Average

The Autoregressive Integrated Moving Average (ARIMA) model is a popular forecasting method used for sales prediction and forecasting. It is a versatile model that can handle time series data with trends, seasonality, and irregular fluctuations.
Methodology:
Autoregressive (AR) Component: The AR component of the ARIMA model considers the relationship between an observation and a certain number of lagged observations (known as the order of autoregression, denoted as p.
Differencing (I) Component: The differencing component involves taking the differences between consecutive observations to remove trends and make the data stationary d).
Moving Average (MA) Component: The MA component takes into account the relationship between the error terms and the lagged error terms (known as the order of moving average, denoted as q).
Model Identification: The identification of the appropriate values for p, d, and q is crucial in ARIMA modelling order of differencing required.

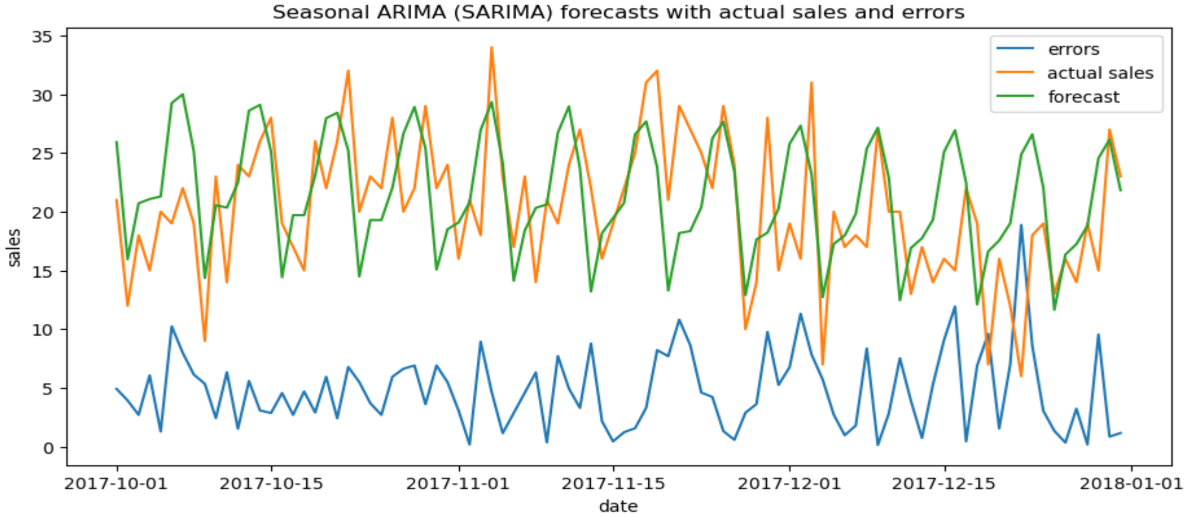
Model Estimation: Once the model order (p, d, q) is determined, the ARIMA model is estimated using methods like maximum likelihood estimation.
Forecasting: The ARIMA model utilizes the estimated coefficients and the historical sales data to generate forecasts for future time periods.

Prediction and Forecast of Sales:



- Analysis of prediction and forecast is done on the test data and forecast is done up to next two years.

Actual Sales and Errors:



Error Data Frame:

model	total_sales	total_pred_sales	SARIMA_overall_error
SARIMA	1861	1973.197449	-112.197449 .

MAE	RMSE	MAPE
4.799933	5.839147	23.72884

4.Supervisied Machine Learning: Linear Regression

The Linear Regression model is a supervised machine learning algorithm widely used for sales prediction and forecasting. It aims to establish a linear relationship between the independent variables (features) and the dependent variable (sales) to make predictions and forecast future sales based on the given input data.

Methodology:

Feature Selection and Engineering: Identify the relevant features that may have a significant impact on sales. This step involves analyzing the correlation between each feature and sales and selecting the most informative ones.

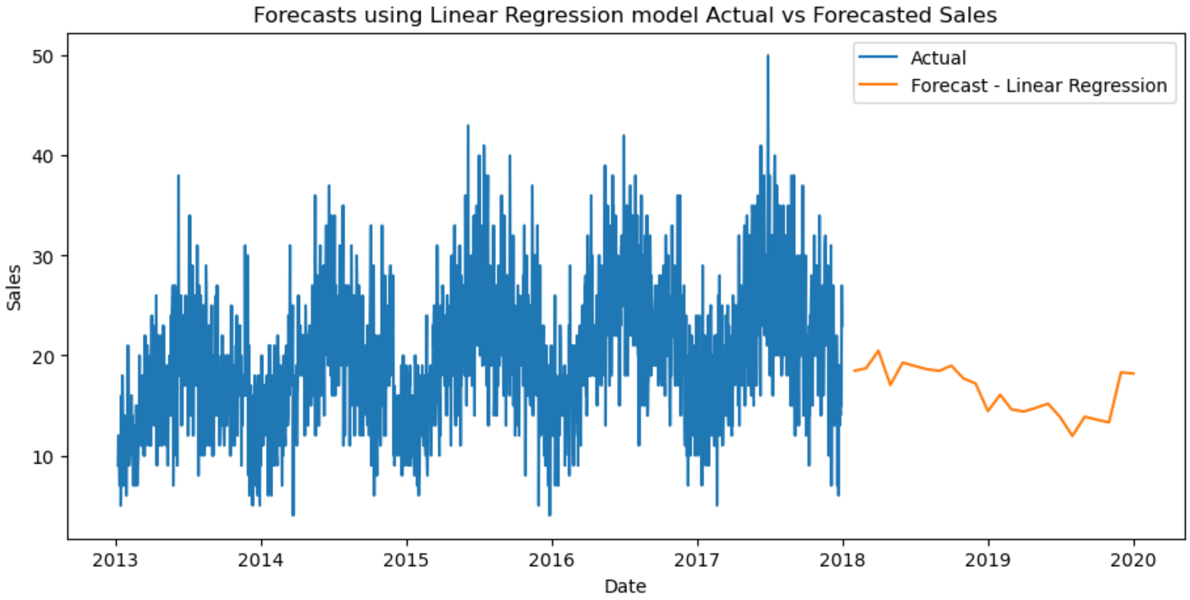
Model Training: Fit the Linear Regression model using the training data. The model will learn the coefficients (weights) associated with each feature to establish a linear relationship between the independent variables and the dependent variable.

Model Evaluation: Evaluate the performance of the trained model using the testing data. Common evaluation metrics for regression models include Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), or R-squared (coefficient of determination).

Model Deployment and Forecasting: Once the model is deemed satisfactory, it can be deployed to predict and forecast sales for new or future data points.

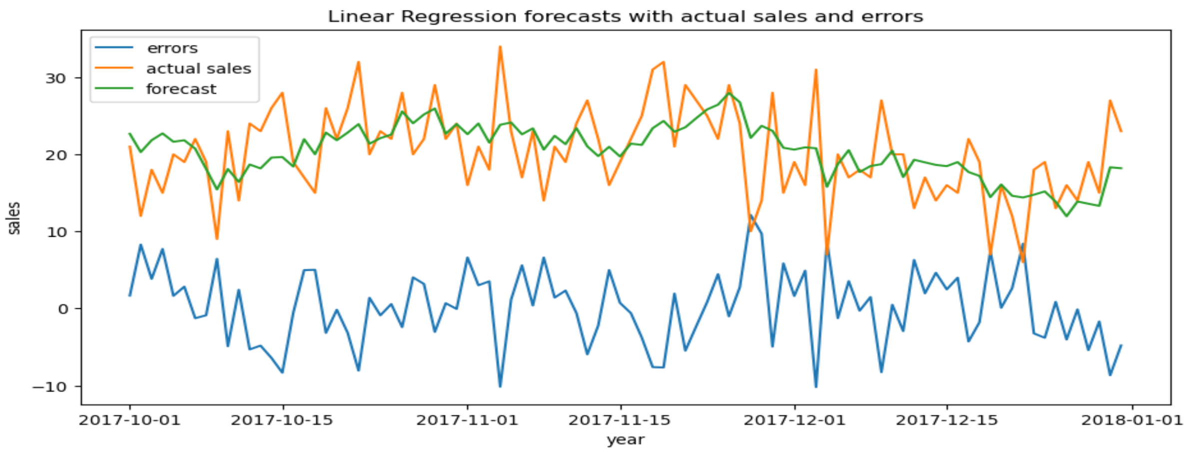
Model Monitoring and Iteration: Continuously monitor the model's performance and retrain the model periodically using updated data. This allows the model to adapt to changing trends and ensure its predictions remain accurate over time.

Prediction and Forecast of Sales:



- Analysis of prediction and forecast is done on the test data and forecast is done up to next two years.

Actual Sales and Errors:



Error Data Frame:

model	total_sales	total_pred_sales	**LR**_overall_error	
LinearRegression	1861	1882.074831	21.074831	-.

MAE	RMSE	MAPE
3.858646	4.786183	19.075519

5.Random Forest Model

Random Forest is a popular machine learning algorithm used for sales prediction and forecasting. It is an ensemble learning method that combines multiple decision trees to make predictions and generate forecasts. The Random Forest model is known for its robustness, accuracy, and ability to handle complex relationships in the data.

Methodology:

Feature Selection and Engineering: Identify the relevant features that may have a significant impact on sales. This step involves analyzing the importance of each feature using techniques like feature importance or correlation analysis.

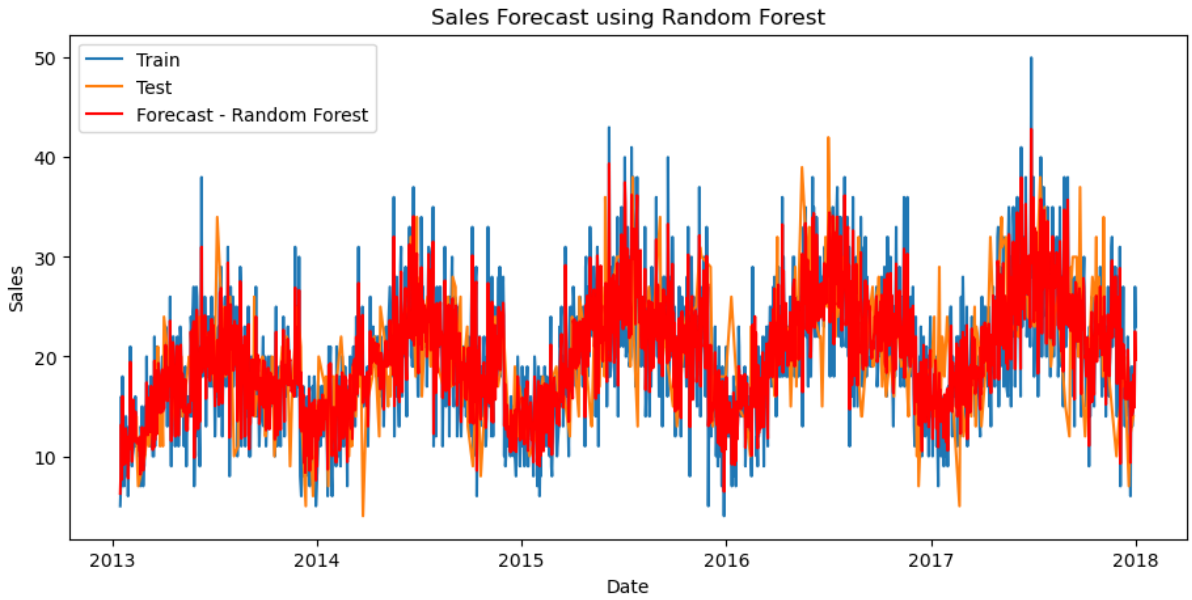
Model Training: Train the Random Forest model using the training data. The model creates an ensemble of decision trees, where each tree is trained on a random subset of the data and a random subset of features. **Model Evaluation:** Evaluate the performance of the trained Random Forest model using the testing data. Common evaluation metrics for regression tasks include Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), or R-squared (coefficient of determination).

Feature Importance: Random Forest models provide feature importance rankings, which indicate the relative importance of each feature in predicting sales. This information helps identify the key factors that significantly influence sales.

Model Deployment and Forecasting: Once the Random Forest model is deemed satisfactory, it can be deployed to predict and forecast sales for new or future data points.

Model Monitoring and Retraining: Continuously monitor the model's performance and retrain the model periodically using updated data. This allows the model to adapt to changing patterns and ensure its predictions remain accurate over time.

Prediction and Forecast of Sales:



Actual Sales and Errors:

Error Data Frame:

model	total_sales	total_pred_sales	**RF_**overall_error	
RandomForest	7272	7336.716667	64.716667	-.

MAE	RMSE	MAPE
3.967006	5.005975	23.669701

Comparison of the Models:

Based on the Error Data Frame:

Among the models evaluated, the Linear Regression model achieved the lowest overall error of 21.074831, indicating better accuracy in sales prediction. It also exhibited the lowest MAE and RMSE values, suggesting better precision in estimating sales values. The SARIMA model performed reasonably well with an overall error of -112.197449 and exhibited a lower MAPE, indicating a good fit to the data.

It is crucial to consider that the choice of the most suitable model depends on several factors, such as the specific requirements of the sales prediction task, the characteristics of the data, and the importance placed on different evaluation metrics. Further analysis and consideration of these factors are essential to make an informed decision regarding the most accurate model for sales prediction.

Based on the comparison, the models can be ranked as follows:

1. **Linear Regression Model:** This model achieved the lowest overall error, MAE, and RMSE values, indicating better accuracy and precision in sales prediction. It also had a relatively low MAPE, suggesting a good fit to the data.

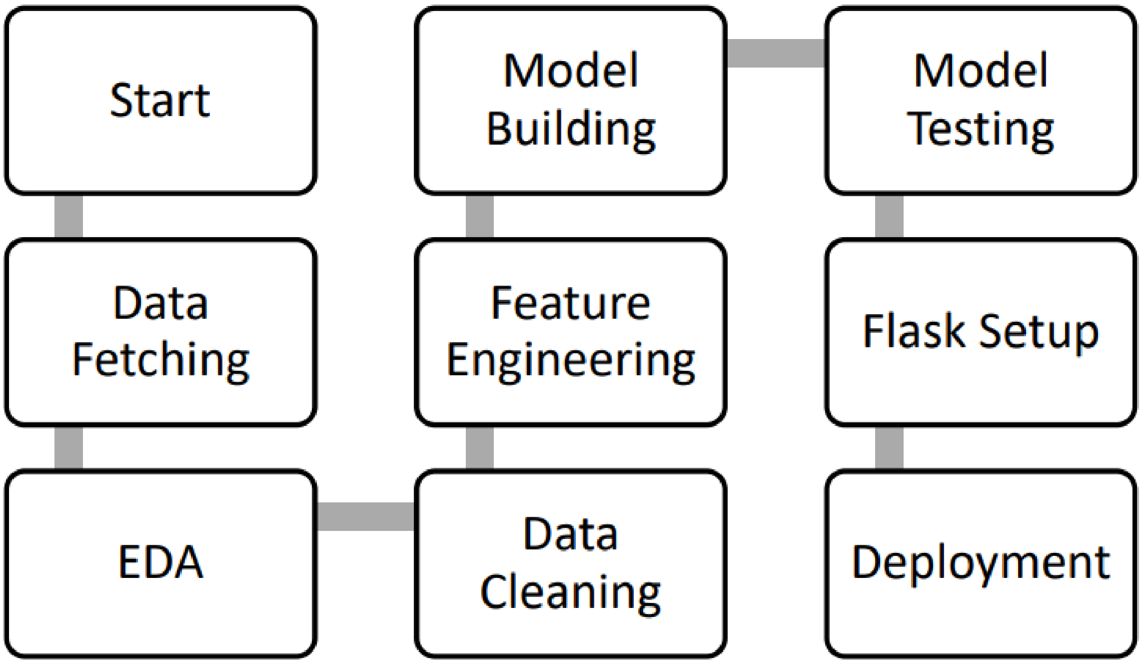
2. **SARIMA Model:** The SARIMA model performed well with a negative overall error, indicating that it slightly overestimated the sales. It also had a relatively low MAE, RMSE, and MAPE, demonstrating good accuracy and precision in sales prediction.

3. **Holt-Winters Models:** Both Holt-Winters models exhibited higher overall errors, MAE, RMSE, and MAPE compared to the top models, indicating lower accuracy and precision in sales prediction. The model with damping performed slightly better than the model without damping.
4. **Seasonal Naive Model:** This model had a moderate overall error and showed reasonable accuracy in sales prediction. However, it had slightly higher MAE, RMSE, and MAPE compared to the top-performing models.

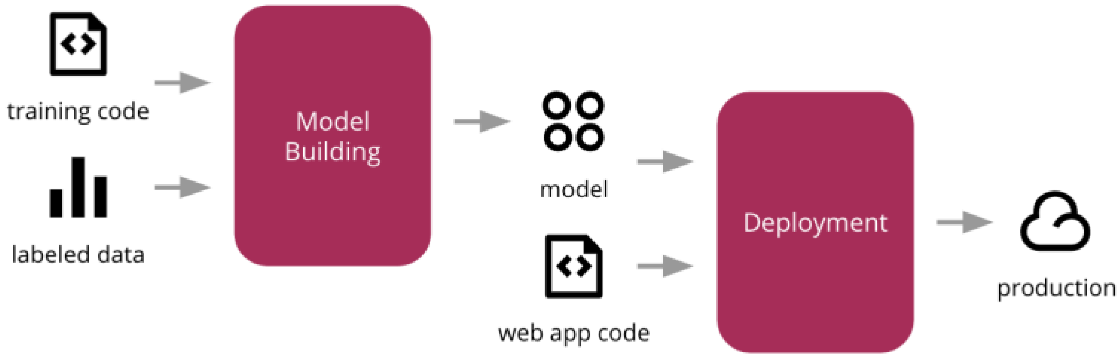
5. **Random Forest Model:** The Random Forest model had the highest overall error, indicating lower accuracy in sales prediction. However, it had a relatively low MAE and RMSE compared to some other models, suggesting reasonable precision.

It is important to note that the selection of the most suitable model depends on the specific requirements of the sales prediction task and the importance placed on different evaluation metrics. It is to consider these factors and conduct further analysis or experimentation to determine the best model for the given scenario.

. Process Flow



Model Training and Evaluation



FRONT END:

I. Generating the model

-using the data set from Kaggle.

1. Identifying target and independent features

-From the dataframe, observe the target column and rest of the columns are independent features

2. Cleaning the data set

-ensure there are no null values in the data.

3. Exploratory Data Analysis

-feature selection steps.

4. Building a model

- regression model for our dataset since it contains a lot of categorical features. This skips the step of label encoding categorical features since can work on categorical features directly.

5. Check model accuracy

- create true predictions from test dataset.

-run the trained model on test dataset to get model predictions and check model accuracy

6. Save the model into a pickle file

- save our model into a pickle file and then save it.

BACK END:

II. Creating backend API from model

-use Python_Flask to create our backend APIs.

- To install Flask pip package: pip install -U Flask

Code to create a basic API:

```
from flask import Flask, jsonify, make_response, request, abort
app = Flask(__name__)
@app.route("/")
def hello():
return "Hello World!"
if__name___== "__main__":
app.run()
```

III. Deploying backend API to Heroku

-to deploy it on a remote server to be used from elsewhere.

Heroku: Heroku an API hosting platform to deploy it on a remote server to be used from elsewhere.

IV. Creating a client-side app using react and bootstrap

- Node.js installed and set-up properly on our machine. So, download and install Node.js for your relevant OS and system.

V. Deploying the client-side app

-deploying client-side app by enhancing security.

Model Efficiency

- Improve the model predictions by adding more data to the model.
- We can use a stack of combined models to improve model efficiency a bit further.
- This cycle of steps will be continued until a certain date arrives.

13. CONCLUSION:

Sales forecasting is a complex yet essential component of business intelligence—like purchasing and budgeting. It helps companies manage their cash flow and identify ways to improve their returns.

Sales forecasting requires extensive machine learning and statistics knowledge.

Predictive analytics and machine learning algorithms have become an essential part of how companies forecast sales. They can be used in conjunction with traditional methods or alone. Either way, the goal is to accurately predict how much sales a company can expect during a given period based on past sales, demographic trends, and behavioral indicators.

In conclusion, no-code predictive analytics is a game-changer for any business looking to gain an edge in their sales forecasting. By leveraging the power of machine learning, businesses can now make data-driven decisions with confidence, identify market trends, and stay ahead of the curve.

As the market becomes increasingly competitive and unpredictable, no-code predictive analytics is quickly becoming a must-have for businesses that want to thrive. The technology has the potential to transform sales forecasting from an art to a science, making it easier, more accurate, and less time-consuming. With the right no-code predictive analytics, businesses can gain a clear understanding of their sales trends, anticipate future shifts, and make informed decisions that drive growth.

GitHub Links:

https://github.com/moulimetturu/Big-Market-Sales-Prediction

https://github.com/BhodigamAkshitha/sales-analysis/blob/05877b8996aaa06976ca498d4590a7ebfac96d33/Sales%20Prediction%20and%20Forecast%20Analysis.ipynb

Business Modelling:

Deploying ML models for small businesses for free (at least initially) will help us gain access to data and check the viability of our model(s). This will also help us to get access to more data which can be used for further improvement of the model then we can implement the pay-per-use, subscription-based models, or tiered pricing based on usage or features. By incorporating business modelling into sales forecasting, small businesses can gain valuable insights, make informed decisions, and increase the accuracy of their sales projections.

Financial Modelling:

In sales forecasting, a financial model equation helps estimate future sales based on various factors and assumptions. While the specific equation may vary depending on the business and industry, here is a general equation that can be used as a starting point:

Projected Sales = (Historical Sales) x (Growth Rate)

In this equation:

- "Projected Sales" represents the estimated sales for a specific future period.
- "Historical Sales" refers to the past sales data, typically derived from historical financial records.
- "Growth Rate" is the assumed rate at which sales are expected to increase or decrease in the future. It is usually expressed as a percentage.

This simple equation assumes a linear growth rate, which means it assumes sales will increase or decrease at a constant rate over time. However, in practice, sales growth rates may vary and are often influenced by multiple factors. Therefore, it's important to consider additional variables and refine the equation to make it more accurate.

Here are a few examples of how the equation can be expanded to incorporate other factors:

1. Multiple Growth Drivers: Instead of using a single growth rate, you can incorporate multiple drivers that impact sales growth. For example:

Projected Sales = (Historical Sales) x (Growth Rate 1) x (Growth Rate 2) x ...

Growth Rate 1, Growth Rate 2, etc., can represent different factors like market growth, customer acquisition rate, or product expansion.

2. Seasonality: If your business experiences seasonal fluctuations in sales, you can adjust the equation to include seasonal factors. For example:

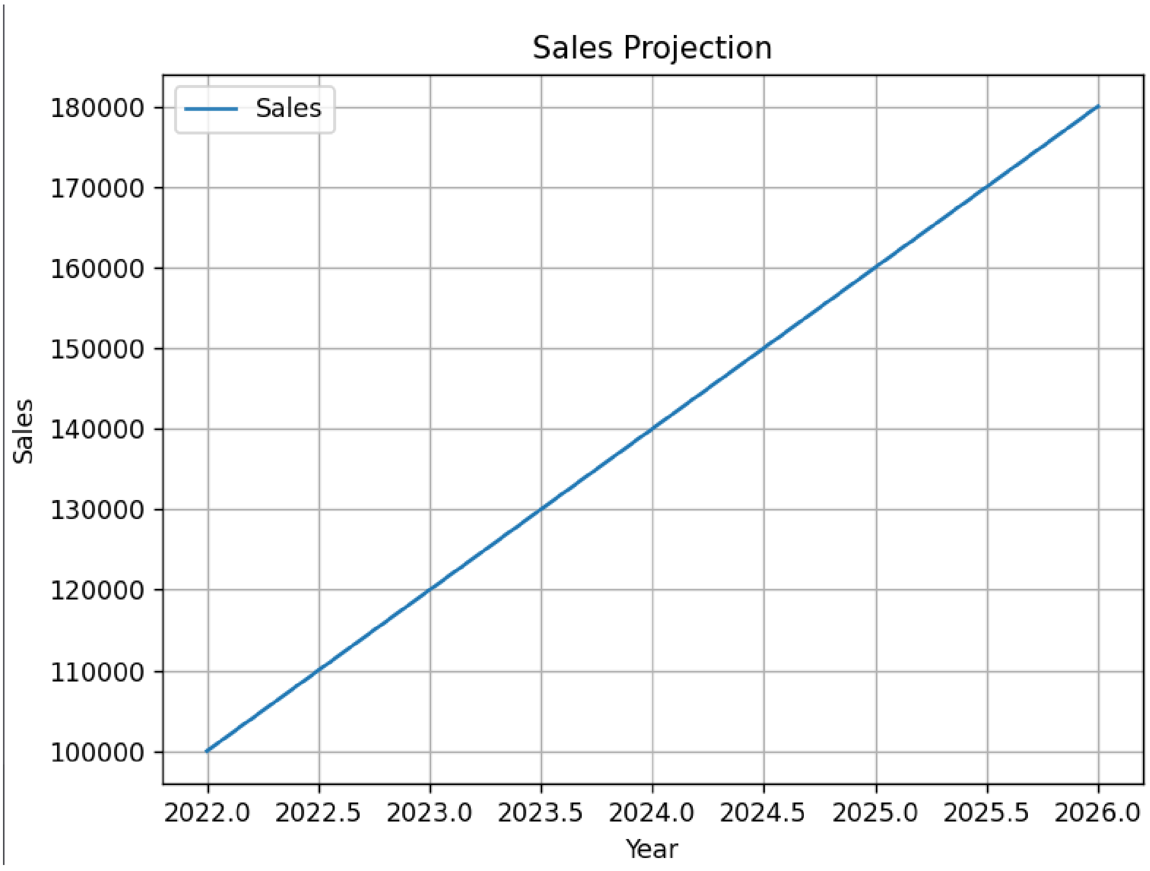
Projected Sales = (Historical Sales in Same Season) x (Seasonal Growth Rate)

This equation compares sales during the same season in previous years and applies a seasonal growth rate to estimate future sales.

3. Product or Segment-Based Forecasting: If you have different products or customer segments, you can create separate equations for each. For example:

Projected Sales (Product A) = (Historical Sales (Product A)) x (Growth Rate

(Product A))
Projected Sales (Segment X) = (Historical Sales (Segment X)) x (Growth Rate (Segment X))
This allows for more specific sales forecasting tailored to individual products or segments.
The specific equation and variables used in the financial model will depend on the nature of your business, available data, and industry dynamics. It's important to continuously refine and update the equation based on new information and feedback to improve the accuracy of your sales forecasts.



See a line plot representing your self-created data.

Conclusion:

Machine learning is a powerful tool for improving sales forecasting in various ways. By analyzing large amounts of data, including historical sales data, customer demographics, market trends, and more, machine learning algorithms can identify patterns and make accurate predictions.
Some of the key applications of machine learning in sales forecasting include demand forecasting, customer segmentation, lead scoring, price optimization, sales pipeline analysis, and sales performance analysis.
Machine learning models can predict future demand for products or services, enabling businesses to optimize inventory management and resource allocation. By segmenting customers based on their preferences and behaviors, businesses can tailor their sales and marketing strategies for different customer groups. Lead scoring algorithms help prioritize leads with higher conversion potential, improving sales efficiency. Machine learning can also assist in determining optimal pricing strategies based on market dynamics and customer behavior. Additionally, machine learning models can analyze sales pipeline data to provide insights into deal probabilities and sales performance. These insights help sales teams prioritize efforts and make data-driven decisions.
To leverage machine learning in sales forecasting, businesses need access to quality data, expertise in machine learning techniques, and suitable tools or platforms. Small businesses can explore cloud-based machine learning services or collaborate with data scientists to incorporate machine learning into their forecasting processes.
Overall, machine learning empowers small businesses to enhance their sales forecasting accuracy, optimize decision-making, and improve overall sales performance.