UNIVERSITY OF SUNDERLAND

DATA SCIENCE: DEMAND ANALYTICS. PREDICTIVE ANALYTICS HELP FORECAST FUTURE DEMANDS ON VARIOUS LEVELS WITH THE HELP OF CURRENT SALES

STUDENT NAME: ABDULAZEEZ OLADIPUPO ADIGUN
STUDENT NUMBER: 219143011

1.1 Introduction

It is possible to utilize predictive analytics to generate a number of predictions regarding events that have yet to take place. From data mining and statistical analysis to data mining, modelling, and artificial intelligence, they carry out their responsibilities. Based on current data analysis, organizations may estimate their future performance using predictive analytics. Analytical methodologies bring together the management, IT, and business modelling operations of an organization in order to make predictions about the future of its business. Data from past and current transactions might also be utilized by PA (Predictive Analysis) to anticipate potential threats and opportunities in the future. Risks are analyzed and scored by PA based on the various operational patterns of a company. Analyzing data in this manner might help a firm acquire a better sense of its future productivity. Here, we'll talk about how to use predictive analytics in real-world situations (Güven, 2021).

1.2 Agile Development in Data science

It is difficult to define what is meant by "agile software development" because there are so many different ways. Using Bhatti's definition, Agile is an extension of the scientific method that emphasizes organized procedures that are based on hypotheses, learning, and observation (2017). If social media usage is any indication, the term "data science" is all the rage among academics. It's a fast-moving field with a lot of potential. Data science has allowed a wide range of sectors to take advantage of new business prospects. To enhance customer service and product quality, businesses use data science to make changes to services and processes that can be implemented in the real world. Just a few examples of agile methodologies are Scrum, Crystal, and Lean Development. Implementing a feature-driven approach is one of these methods. Every method has a special way of approaching solving a problem. There is one element that all techniques have in common with the agile principles: a shared attribute. In agile methods, iteration and feedback are used to improve the system being developed (Huber, 2020).

Three typical agile development concepts are iteration, feedback, and feature selection. It is possible to use these notions in data science. Software experiments need a hypothesis-formation process similar to that used in feature definition. With test-driven implementations, iterations are very much the same. Principles like these are essential in data science (Wiyanti, 2021).

1.3 Product Design Section

1.3.1 Data Source

I selected the dataset as from the open source website https://data.gov.uk/

To help the public find and use free and open government data and support data publishers since 2010, data.gov.uk was launched. Find open data was added to our site in March of this year. Policymaking and service improvement can both benefit from better data collection. Internal departments can combat and decrease fraud and waste. In order to build trust with our citizens and deliver more cost-effective and personalized services, we must utilize data responsibly, freely, and innovatively. The digital transformation of government is not possible without the use of data in the Transformation Strategy. Without data, we can do nothing about government change. Organizations can publish data on Find Open Data by linking to GOV.UK datasets. To help in the production and dissemination of data on GOV.UK (Ensafi, 2022).

1.3.2 Application

When it comes to business in COVID-19, it's all about human behavior, which causes quick swings in the market. New experiences and epidemic limits can both influence human behavior. Machine learning techniques may be used to estimate demand and predict future sales in this article. On the subject of the COVID-19 pandemic, I'll discuss ways to improve forecast accuracy. The retail company will serve as an example because I've worked on forecasting models for retail field items before. Don't be concerned if your company's primary concentration is not retail. This article's primary purpose is to explain how machine learning may be used to estimate demand in both steady and crisis situations (Feizabadi, 2020).

1.3.3 Specification and Functional and Non-Functional Requirements

Features that contribute to the business's success are known as "agile features." Many methods exist for incorporating new or modified functionality into features. If a credible estimate of the data science work is to be developed, the development team must base its feature definition on the business value. Iteration compatibility is a need for data science features. An important consideration when using agile techniques in scientific research is to guarantee that the outcome is marketable for potential clients. The deliverable functionality that must be satisfied is specified by the features established in data science. Variable client engagement and project stages are among the benefits (Tekin, 2021).

1. **Functional requirements**

These are the details of how the product should function. In a nutshell, functional requirements describe what a piece of software should do and how it should respond to user inputs. Basically, if you fail to fulfil the software's functional requirements, it will not work. When working on software projects, business analysts develop both functional and nonfunctional requirements. Effective communication with clients is essential to ensure that their needs and wishes are being met (Bhutti, 2017).

2. **Non-Functional requirements**

Even if the system's essential behavior is determined by its functional requirements, the non-functional requirements explain how this function will be carried out. We'll come back to it in the email notification example later on in this post. The system must be able to send out email notifications on its own to be functional. Emails should be distributed based on non-functional requirements (Özalp, 2021).

1.3.4 Architecture design

Decision-making and resource allocation are made better in nearly any organization when reliable future forecasts are available. By successfully forecasting surges in demand for products and services, businesses may gain a competitive advantage. In order to effectively respond to fluctuations in demand and avoid excessive stockpiling, they have improved their forecasting capabilities. Only a few examples are given here: retail and internet shops, hospitals, and power grids (Cao, 2017).

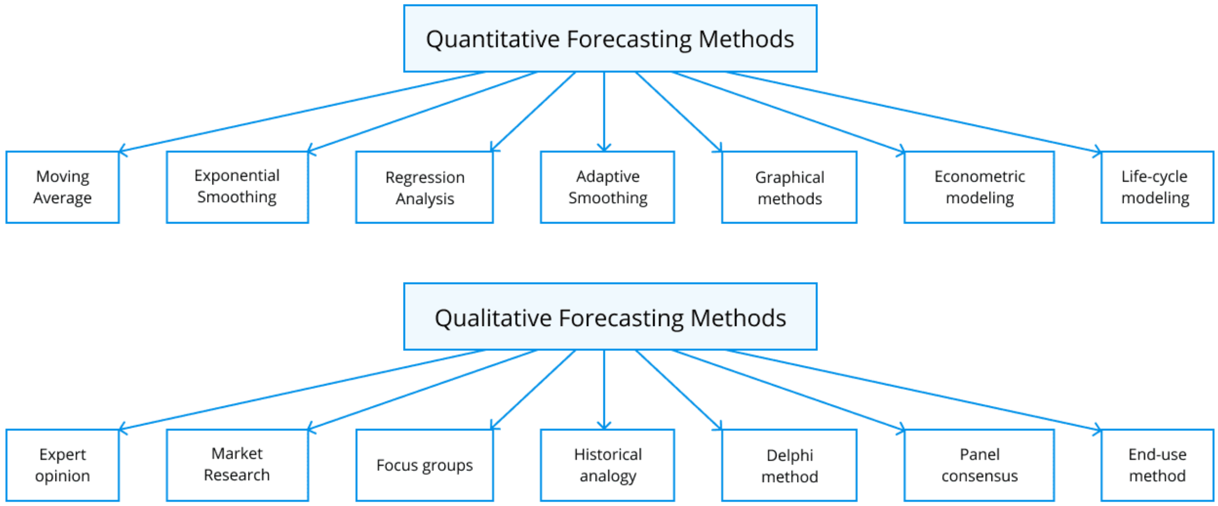


Fig 1.1 Demand Forecasting ML Methods

In order for a company to be profitable and sustainable, it must be able to effectively forecast demand for its products. The marketing staff's empirical knowledge, however, can lead to substantial errors in this process in the majority of cases. We devised a production planning architecture based on demand analysis using business intelligence architecture and analytical algorithms. Using a case study to support the design has resulted in an 85 percent success rate, which is a significant improvement. We feel that the approach described here might be applied to other organizations. Big data and consumer behavior theory are used to construct a demand prediction system for online merchants that aims to increase regional forecasting accuracy while also cutting down on forecasting timeframes. E-commerce enterprises must increase their capacity to predict for both business and customer reasons. Initial Demand, Possible Demand, Core Demand, and Effective Demand are all part of an ML Model-based framework for analysis and design (Tripp, 2018).

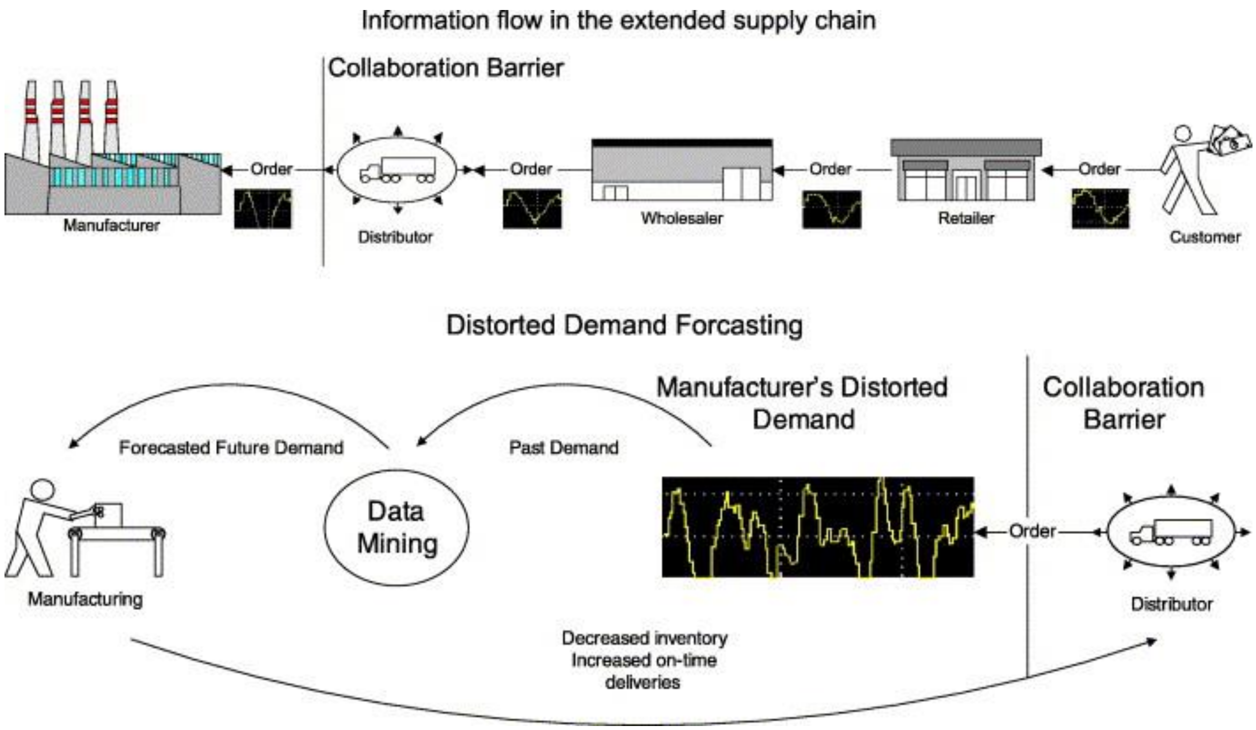


Fig 1.2 Architecture of demand forecasting

1.3.5 Use case specifications

In software development and data science, a user narrative is a critical component of the system. As a consequence of customer feedback, a customized system has been developed. In order to understand, explain, and predict a phenomenon in data science, there is no alternative to user input. In data science, the research question is the closest thing to a user's input. There is a question from the product owner about an incident that has to be handled with the data supplied. To priorities queries from the product owner, it is vital to understand why they asked them in the first place and the possible consequences of resolving them. Consequently, data scientists must provide the answers by converting datasets and creating models to meet the questions. Web services that scale up or a simple blog post might be examples of responses (Kulmp, 2017).

Using data science iteration necessitates the preparation of data tables before they can be properly understood. It is always the twenty-fifth inquiry in a sequence of queries that provides the greatest information. Accordingly, graphs tend to have more informative ones in their third or fourth rows than in their top ones. In order to create a model that is both accurate and acceptable, several iterations of hyper parameters and feature engineering may be required. Analysis of data requires a high degree of repetition, which is essential for visualization, extraction, and the production of insights (Kobielus, 2017).

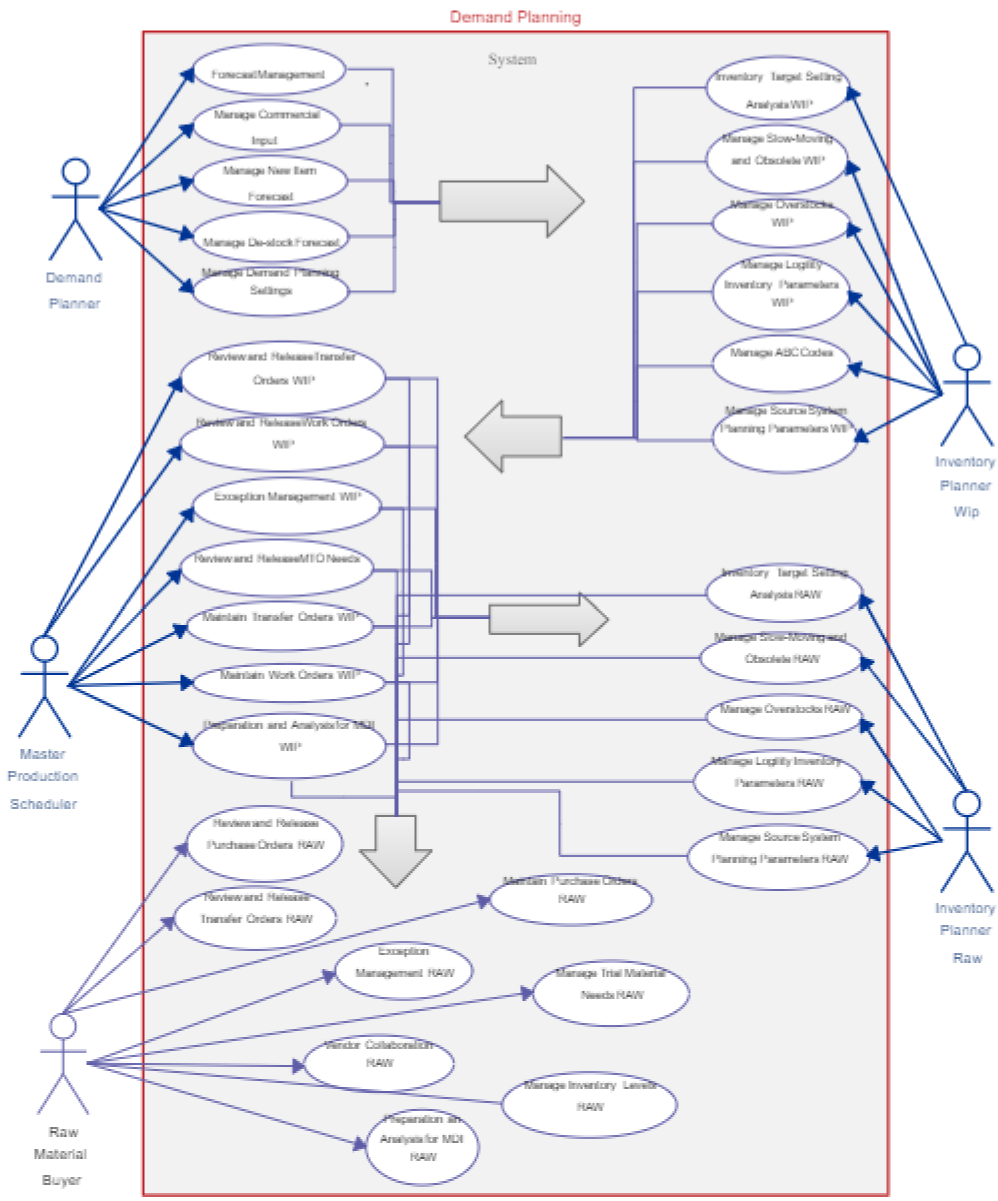


Fig 1.3 Use Case for Demand Forecasting

1.4 Product Development Section

1.4 1 Selection of ML Tool

Colab is the online available google Machine Learning tool which is used in our proposed demand forecasting

The process of predicting future demand for a certain product is known as demand forecasting. When it comes to developing and stocking new items, manufacturers and retailers alike may benefit from this information. In supplier relationship management, you can readily determine whether new supply chains or fewer suppliers are needed by knowing how many things your clients will need in advance and how many of each type you should acquire. Customers who want to purchase a product have a reasonable expectation that it will be available when they do so. In the next few months, you may use demand forecasting to predict which products will be purchased from a certain retail location. Increased customer satisfaction and loyalty are a result of this. Software that predicts demand can help supply chains run more smoothly. There will be no unsold inventory taking up valuable shop space, so customers can be sure they are getting what they ordered. Forecasts, which are routinely used to make modifications in advertising and marketing efforts, might impact their efficacy. It is possible to use powerful machine learning forecasting models to take into account data from marketing. Time series-based demand forecasting, which is part of ERP software, predicts production needs based on how many items will be sold in the end (Nelluta, 2018).

1.4.2 Agile Methodology

In software development, agile is well accepted, but it may also be applied effectively in data science initiatives. Factors in data science projects are quite different from those in software challenges when approaches are used to overcome data science difficulties. Before using agile approaches in data science, there must be an understanding of why they are necessary. Many advantages come with data science, and it looks stunning when used in real-world settings. Despite the fact that projects are managed by specialists, data science is not as enticing as it may be viewed when it comes to fixing difficulties. Data science problem-solving is difficult since most initiatives lack a degree of assurance. In the absence of data from others who have worked on comparable projects, it is highly common for someone to think that they are the only ones who understand the problem. Most of the time, things don't go according to plan. This has an effect on the timing of the project since it might be difficult to come up with a clear deadline for the project (Smirnov, 2021).

1.4.3 System testing method

Agile approaches have been implemented in data science projects because of the problems in project management. When it comes to fine-tuning data science initiatives, agile methodology is an excellent tool. When approaches are used for a data science project, a significant level of creativity is necessary. The utilization of cycles makes it easier and more efficient to operate in an agile environment. In every cycle, there's a chance to learn something new that can be used to improve the project's outcomes and shared with other organizations. Scientists claim that agile techniques improve their productivity by assisting them in optimizing the results of projects (Xue, 2021).

Fig 1.4 Agile Testing

1.4.4 User Evaluation Plan

The performance of the model is evaluated and compared to projected outcomes during this phase. Using data, algorithms may be trained, and these algorithms can then be tested on real-world data to determine their correctness, for example. The parameters of the model were compared to those of the baseline model to determine how well it performed. This gives you ample time to look for flaws in the model, which will give specific ideas on how to enhance it (Sharma, 2019)

1.5 Results

1. **Dataset Preview**

df.head(5)

	Product_Code	Warehouse	Product_Category	Date	Order_Demand
0	Product_0993	Whse_J	Category_028	2012-07-27	100
1	Product_0979	Whse_J	Category_028	2012-01-19	500
2	Product_0979	Whse_J	Category_028	2012-02-03	500
3	Product_0979	Whse_J	Category_028	2012-02-09	500
4	Product_0979	Whse_J	Category_028	2012-03-02	500

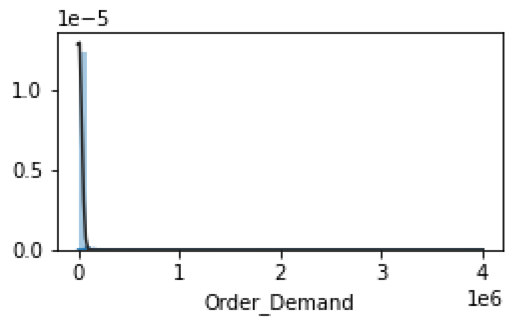
2. **Removing the missing values**

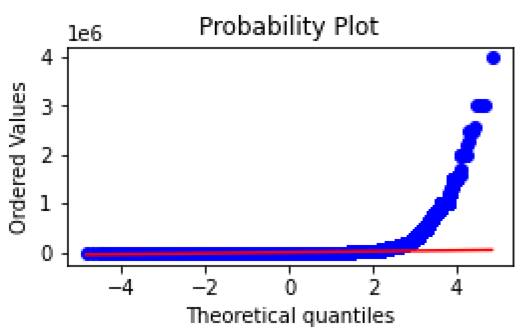
```
df.dropna(axis=0, inplace=True)
df.reset_index(drop=True)
df.sort_values('Date')[10:20]
```

	Product_Code	Warehouse	Product_Category	Date	Order_Demand
75193	Product_0642	Whse_C	Category_019	2011-10-31	3
121820	Product_0202	Whse_A	Category_007	2011-11-04	(100)
121819	Product_0202	Whse_A	Category_007	2011-11-04	(400)
131028	Product_2143	Whse_S	Category_009	2011-11-18	(25)
131031	Product_0131	Whse_S	Category_021	2011-11-18	(12)
131032	Product_0288	Whse_S	Category_021	2011-11-18	(50)
44450	Product_0980	Whse_A	Category_028	2011-11-18	4000
131027	Product_2138	Whse_S	Category_009	2011-11-18	(49)
131026	Product_2137	Whse_S	Category_009	2011-11-18	(25)
44795	Product_0965	Whse_A	Category_006	2011-11-18	1

3. Distribution of Order Demand

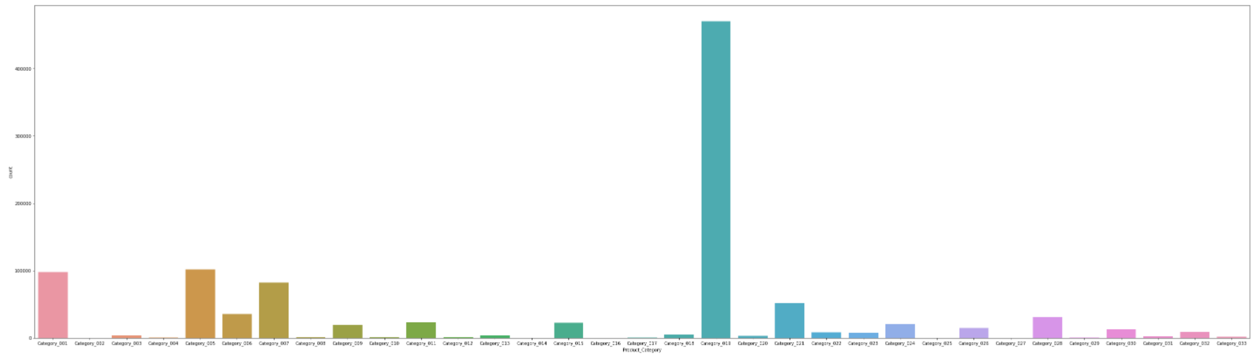
```
rcParams['figure.figsize'] = 4,2

sb.distplot(df['Order_Demand'], fit=norm)
fig = plt.figure()
res = stats.probplot(df['Order_Demand'], plot=plt)
plt.show()
```





4. Product Category

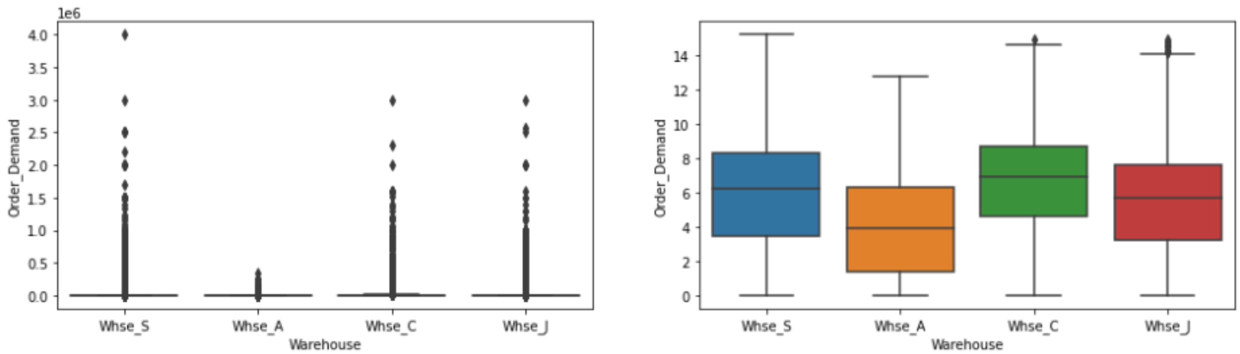
```
print (len(df['Product_Category'].value_counts()))
rcParams['figure.figsize'] = 50,14
sb.countplot(df['Product_Category'].sort_values(ascending = True))
```



5. **Orders Demand by Warehouse**

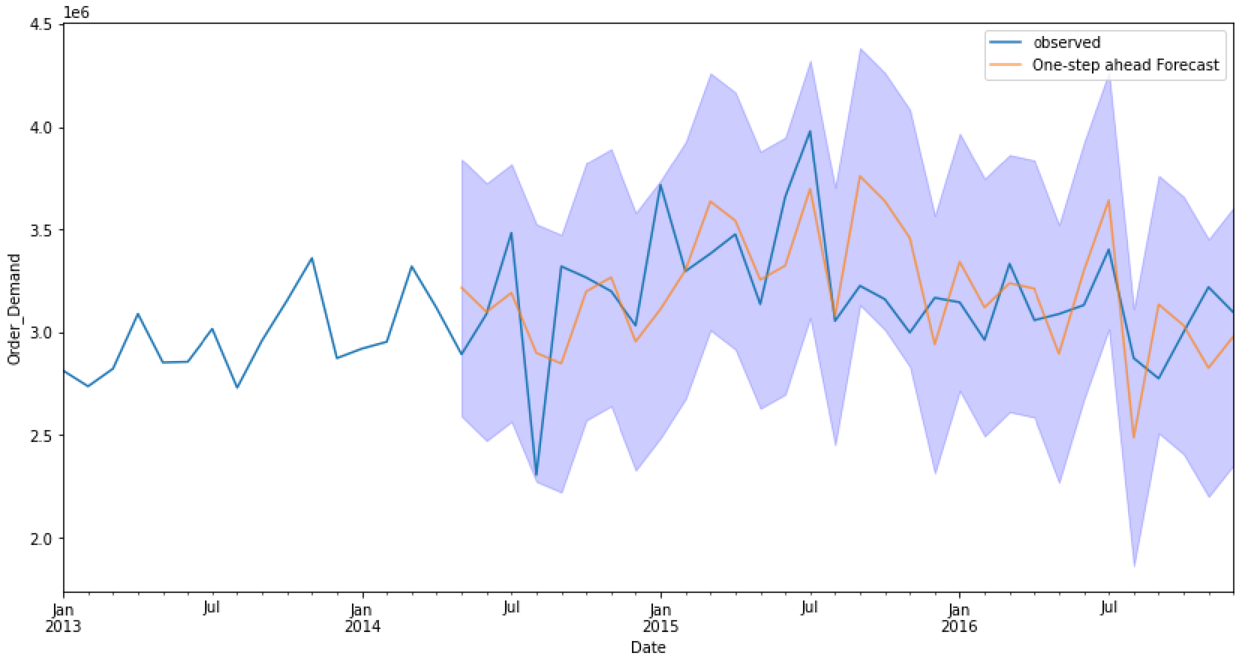
1.6 Project Management Section

```
rcParams['figure.figsize'] = 16,4
f, axes = plt.subplots(1, 2)
fig3 = sb.boxplot( df['Warehouse'],df['Order_Demand'], ax = axes[0])
#Data with Log Transformation
fig4 = sb.boxplot( df['Warehouse'], np.log1p(df['Order_Demand']),ax = axes[1])
del fig3, fig4
```

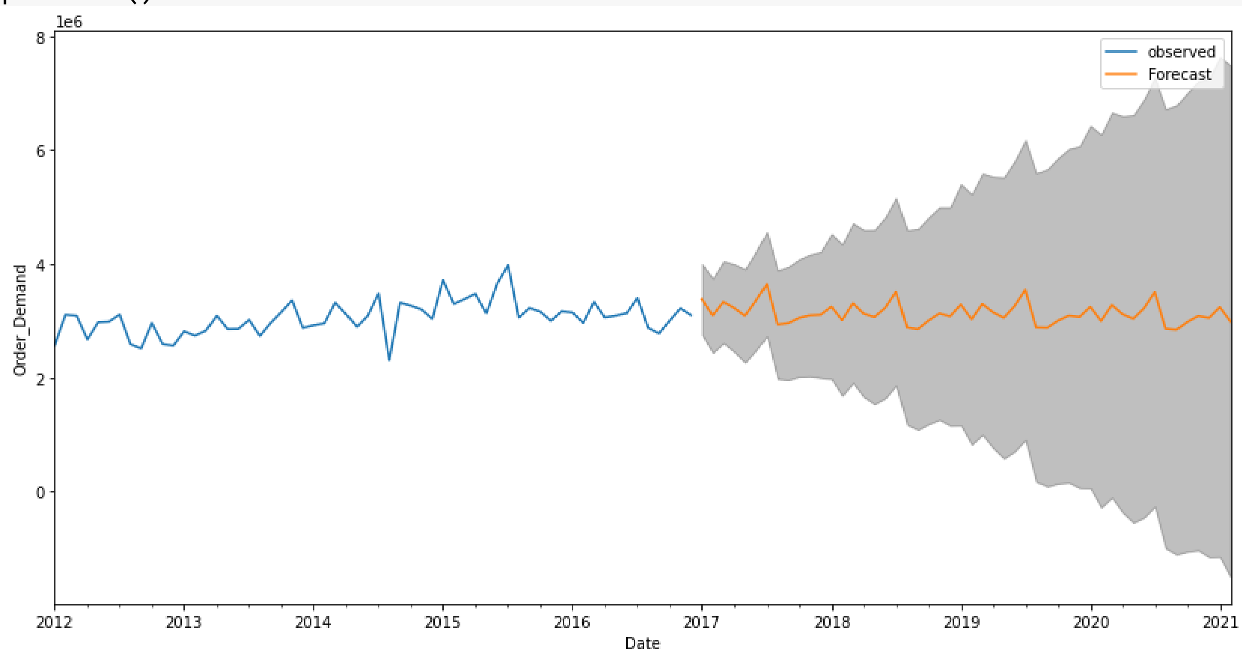


6. **Forecasting Demand**

```
pred = results.get_prediction(start=pd.to_datetime('2014-05-01'), dynamic=False)
#false is when using the entire history.
#Confidence interval.
pred_ci = pred.conf_int()
ax = y['2013':].plot(label='observed')
pred.predicted_mean.plot(ax=ax, label='One-step ahead Forecast', alpha=.7, figsiz
e=(14, 7))
ax.fill_between(pred_ci.index,
                pred_ci.iloc[:, 0],
                pred_ci.iloc[:, 1], color='blue', alpha=.2)
ax.set_xlabel('Date')
ax.set_ylabel('Order_Demand')
plt.legend()
plt.show()
```




```
pred_uc = results.get_forecast(steps=50)
pred_ci = pred_uc.conf_int()
ax = y.plot(label='observed', figsize=(14, 7))
pred_uc.predicted_mean.plot(ax=ax, label='Forecast')
ax.fill_between(pred_ci.index,
                pred_ci.iloc[:, 0],
                pred_ci.iloc[:, 1], color='k', alpha=.25)
ax.set_xlabel('Date')
ax.set_ylabel('Order_Demand')
plt.legend()
plt.show()
```



REFERENCES

Güven, İlker & Uygun, Ozer & Simsir, Fuat. (2021). Machine Learning Algorithms with Intermittent Demand Forecasting: An Application in Retail Apparel with Plenty of Predictors. TEKSTİL VE KONFEKSİYON. 31. 10.32710/tekstilvekonfeksiyon.809867.

Huber, Jakob & Stuckenschmidt, Heiner. (2020). Daily retail demand forecasting using machine learning with emphasis on calendric special days. International Journal of Forecasting. 36. 10.1016/j.ijforecast.2020.02.005.

Wiyanti, Dian & Kharisudin, I & Setiawan, AB & Nugroho, Arief. (2021). Machine-learning algorithm for demand forecasting problem. Journal of Physics: Conference Series. 1918. 042012. 10.1088/1742-6596/1918/4/042012.

Feizabadi, Javad. (2020). Machine learning demand forecasting and supply chain performance. International Journal of Logistics Research and Applications. 1-24. 10.1080/13675567.2020.1803246.

Ensafi, Yasaman & Hassanzadeh Amin, Saman & Zhang, Guoqing & Shah, Bharat. (2022). Time-series forecasting of seasonal items sales using machine learning – A comparative analysis. International Journal of Information Management Data Insights. 2. 100058. 10.1016/j.jjimei.2022.100058.

Tekin, Ahmet & Cebi, Ferhan & Yağcioğlu, Engin. (2021). Demand Forecasting of a Company that Produces by Mass Customization with Machine Learning. 10.1007/978-3-030-85577-2_23.

Bhatti, B. (2017). Agile for Data Science. Retrieved from: https://medium.com/@thebabar/agile-for-data-science-c8f46daa0dd4

Özalp, Mehmet & Akarsu, Özgür. (2021). DEMAND FORECASTING IN RETAIL: A MACHINE LEARNING BASED APPROACH FOR ACCURATE SHORT-TERM FORECASTS.

Cao, L. (2017). Data science: a comprehensive overview. ACM Computing Surveys (CSUR), 50(3), 43. Retrieved from: https://dl.acm.org/citation.cfm?id=3076253

F. Tripp, J., & Armstrong, D. J. (2018). Agile methodologies: organizational adoption motives, tailoring, and performance. Journal of Computer Information Systems, 58(2), 170-179. Retrieved from: https://www.tandfonline.com/doi/abs/10.1080/08874417.2016.1220240

Klump, J. F., & Robertson, J. (2017, December). Towards seamless workflows in agile data science. In AGU Fall Meeting Abstracts. Retrieved from: http://adsabs.harvard.edu/abs/2017AGUFMIN32A.09K

Kobielus, J. (2017). Agile Development in Team Data Science. Retrieved from: https://wikibon.com/agile-development-in-team-data-sci

Nellutla, V. (2018). Applying Agile IT Methodology to Data Science Projects. Retrieved from: https://www.datasciencecentral.com/profiles/blogs/applying-agile-it-methodology-todata-science-projects

Smirnov, P & Sudakov, Vladimir. (2021). Forecasting new product demand using machine learning. Journal of Physics: Conference Series. 1925. 012033. 10.1088/1742-6596/1925/1/012033.

Zhenchang X, Xue S, Wu L and Sun J 2020 ForeXGBoost: passenger car sales prediction based on XGBoost. Distrib. Parallel Dat. 38 713 https://doi.org/10.1007/s10619-020-07294-y

Sharma S K, Chakraborti S and Jha T 2019 Analysis of book sales prediction at Amazon marketplace in India: a machine learning