# Data Visualization- Food Demand Forecasting Datasets

Anish Paudyal
*Department of Computer Science and Engineering*
*Kathmandu University*
Dhulikhel, Nepal
ap02409818@student.ku.edu.np

Bibek Palanchoke
*Department of Computer Science and Engineering*
*Kathmandu Univerisity*
Dhulikhel, Nepal
bp02409818@student.ku.edu.np

***Abstract*— Here and now, we are in the era of big data. With continuous internet development, huge data is being generated day to day which cannot be observed at any time. Patterns, trends and correlations that might go undetected in text-based data can be exposed and recognized easier with the data visualization techniques. Data visualization turns small as well as large datasets into visual forms which are easier to process and understand for the human brain. This report summarizes the concept of data visualization on the food demand forecasting datasets elaborating and analyzing the visual forms of data using the Streamlit framework. It explains how the users will be able to interact with the information, based on an initial set of hints, charts, tables and reports, produced by the Dashboard itself.**

***Keywords - Data visualization, Datasets, Dashboard, Framework, Streamlit***

## I. INTRODUCTION

With the development of internet, data is rapidly expanding and we have entered the data era. However, the data era is currently in explosive period and we are facing this unprecedented challenge. Life data has become has become the main reason for the rapid growth of the data. Only by analyzing the real data generated, the researchers can predict and summarize the trends of the related fields contained in them based on the data. Traditional data statistics performed by professional data analysts takes a long time and has low effect. Data visualization has become the most efficient way of presenting huge or small amount of data in a interpreted, analytical and understandable manner.[1]

Data visualization is a general term that describes any effort to help people understand the significance of data by placing it in a visual context. Data visualization is the presentation of quantitative information in a graphical form. Data visualizations are surprisingly common in our everyday life, but they often appear in the form of well-known charts and graphs. [2]

It can be used to discover unknown facts and trends. Good data visualizations are created when communication, data science, and design collide. Data visualizations done right offer key insights into complicated data-sets in ways that are meaningful and intuitive. With the goal of making data more accessible and understandable, data visualization in the form of dashboards is the go-to tool for many businesses to analyze and share information.[1]

When you think of data visualization, your first thought probably immediately goes to simple bar graphs or pie charts. While these may be an integral part of visualizing data and a common baseline for many data graphics, the right visualization must be paired with the right set of information. Simple graphs are only the tip of the iceberg. There's a whole selection of visualization methods to present data in effective and interesting ways [3]. The visualization of data includes different types of charts, tables, graphs, geospatial, infographic, dashboards etc. Specifically, most come types of visualizations are area map, bar chart, box plot, whisker plot, bullet graph, gantt chart, heat map, highlight table, histogram, pie chart, tree map and so on.

The following report elaborates in brief about different types of visualization with reference of datasets namely "Food Demand Forecast". This report explains about the importance of visualization of data on huge amount of data for somehow accurate forecasting and prediction and shows how building a dashboard to visualize data is more efficient and applicable than viewing textual tabulated data.

## II. TECHNICAL BACKGROUND

In visualization, data, information and knowledge are three terms used extensively, often in an interrelated context. In many cases, they are used to indicate different levels of abstraction, understanding or truthfulness [4]. Due to the massive amount of information that exists and is created every day, the number of existing artifacts and the need of all users and consumers of knowledge that exist, it is difficult to find ways to present the information in an accessible way or make sense out of it [5].

For the visualization of our datasets, we have used one of the currently available tools known as dashboard. A data visualization dashboard is an interactive dashboard that allows you to track key metrics across multiple marketing channels, visualize the data points and create reports for efficient understanding of given data [6]. Dashboards empower both technical and non-technical users to understand and leverage business intelligence to make more informed decisions. Users actively participate in the analytic process by compiling data and visualizing trends or occurrences, and uncovering an objective view of performance metrics that can be immediately understood [7].

A dashboard is a sort of graphical user interface that frequently enables quick access to key performance indicators (KPIs) related to a certain goal or business activity [8]. There are number of frameworks available for building a dashboard. Here, we have used the framework "Streamlit". Streamlit is a free, open-source, all-Python framework that

allows data scientists to easily construct interactive dashboards and machine learning web apps with no prior knowledge in front-end web development [9]. Streamlit's open-source app framework is the easiest way for data scientists and machine learning engineers to create beautiful, performant apps in only a few hours [10]. Streamlit is a free toolset for creating innovative dashboards. It has a diverse set of UI components. It includes practically every standard user interface component, such as a checkbox, slider, collapsible sidebar, radio buttons, file upload, progress bar, and so on. Furthermore, these components are quite simple to utilise [11]. Particularly, the dashboard built shows and helps to predict and forecast the demand of the food based on the provided sets of data with the help of streamlit.

## III. RELATED WORKS

### A. *Alloy*

Canadian startup Alloy develops planning and execution software to forecast customer demand. Their PoS-driven solution selects forecast models and integrates them with sell-through data for different stock-keeping units (SKUs) or categories. The model then identifies deviations between the forecast and consumer demand in real-time. This data is later analyzed to create responses through supply chains. The solution helps the food industry achieve high levels of supply chain efficiency while minimizing food waste. [12]

### B. *Digitory*

Indian startup Digitory specializes in artificial intelligence (AI) and machine learning (ML) platforms to manage costs and profitability for the food industry. Their sales forecasting solution uses advanced computational techniques to identify sales pattern and create accurate models for prediction. Their solution increases the profits of restaurants through the optimization of production and inventory planning. [13]

### C. *PredictHQ*

United States based startup PredictHQ develops demand forecasting solutions that use internal and external data to learn and adapt its models. Their demand intelligence tool provides grocery stores, supermarkets and restaurants with analysis on events that impact inventory levels, assortment planning and worker schedules. PredictHQ's application programming interface (API) also uses external intelligence such as sports games viewership, nearby expos or community events, weather and more to anticipate event-driven demand volatility for quick-service restaurants. [14]

### D. *Delicious Data*

German startup Delicious Data develops forecasting solutions using machine learning for procurement and production. Their solution helps bakeries forecast sales to enable demand-oriented action and optimize product display placement to attract customers. The software forecasts sales at a minute level which allows employees to optimize the production of products with high accuracy. Thereby, reducing overproduction and wastage of food which, in turn, increases profits. [15]

## IV. PROPOSED SYSTEM/ METHODOLOGY

The purpose of this project is to visualize the datasets. We have taken the datasets namely "Food Demanding Forecasting" from kaggle datasets.

"A forecast is a statement of what is expected to happen in the future, especially in relation to a particular event or situation." (Collins dictionary) Thus, forecasting is a discipline which finds its applications in a high number of different fields. It can be used to predict from weather to volume of phone calls, from road traffic to products' demand. The best practice in each of these fields is specific and, depending on its characteristics, can be extremely straight forward or more complex. [16]

After finding the dataset we researched and studied to determine an appropriate framework to build a dashboard. There are multiple frameworks like R Shiny, Streamlit, Dash by Plotly etc. Out of all the methods to build our framework, we implemented the Streamlit framework. It is an open source framework of python which we used to visualize our datasets.

### A. *Research and Study*

## V. DATASETS

A dataset is an ordered set or collection of data. This set is normally presented in a tabular pattern. It could include information such as facts, numbers, figures, names or even basic descriptions of objects. Datasets can be of type numerical, bi-variate, multivariate, categorical or correlation. For our study, data can be organized in the forms of graphs, charts, tables etc to assist in the analysis of gathered data [17].

In this project, out of all the kaggle datasets, we found the "Food Demand Forecasting" Datasets. The datasets consist of three CSV files. These include: Weekly Demand Data, Fulfillment Center Information and Meal Information. The datasets are from a meal delivery company which operates in multiple cities. They have various fulfillment centers in these cities for dispatching meal orders to their customers. The data can help these centers with demand forecasting for upcoming weeks so that these centers will plan the stock of raw materials accordingly.

## VI. EXPERIMENTAL ANALYSIS

Firstly, we look at the Weekly Demand Data. Figure 1 shows the bar chart for the weekly demand data.

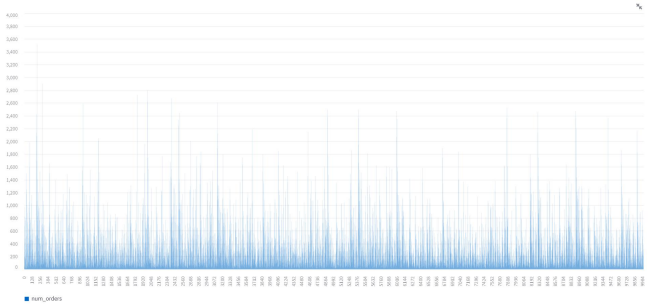

*Figure 1: Bar chart showing number of orders for weekly demand data*

The bar chart shows the number of orders of food in respective weeks. The blue bars in the chart index the number of orders in the weekly demand data. The highest number of orders range from about 3000-3500 whereas the least number of orders range from about 10-20 orders.

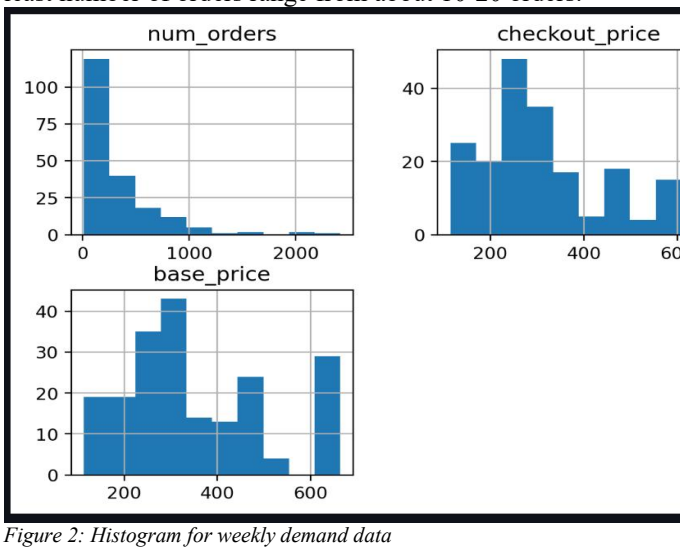


*Figure 2: Histogram for weekly demand data*

Figure 2 shows the plotting of histogram for the weekly demand data. The charts include the histograms of the respective weeks with the columns number of orders, checkout price and base price respectively for the first 200 rows of data. These histogram charts show the frequence of the highest and least frequency of number of orders, checkout and base price for the first 200 rows of data of weekly demand data. The second and third histogram charts show that the checkout price for the given weekly data are h more or less equal to that of base price and in some of the cases relatively higher indicating the profitability in the business.

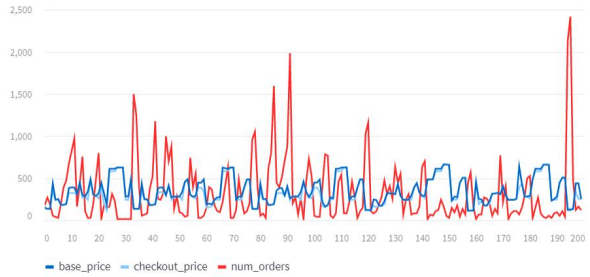


*Figure 3: Line chart for weekly demand data*

Figure 3 represents the line chart for the same dataframe as for previous charts. Here as shown in the indexing, the blue and light blue lines represent checkout and base price respectively. In the line chart, we can see that these two lines are almost overlapping to each other. This indicates that there is not much differences in the checkout and base price for the demanded food. Also, the number of orders show that after a high demand in orders for a range of 10 weeks, the number of orders in the next 100 weeks are usually in decreasing order. If we have to predict the orders of next 10 weeks (i.e., 210-220) the number of orders should be comparatively less than the previous week range when it was very high.

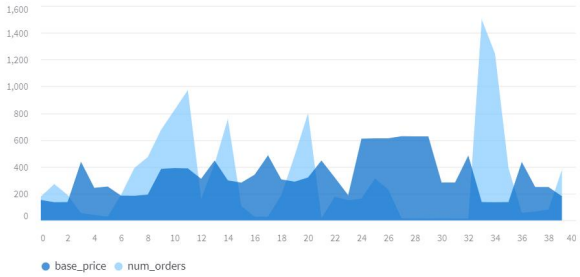


*Figure 4: Area chart for weekly demand data*

Figure 4 shows the area chart for the same dataframe including the columns base price, checkout price and the number of orders for the first 200 rows of data. The area chart combines the line chart and bar chart to show how one or more groups' numeric values change over the progression of a second variable, typically that of time. This chart shows more clearly that base and checkout price almost overlap each other except in some points. It also represents the volume of the columns showing the range or area covered by them.

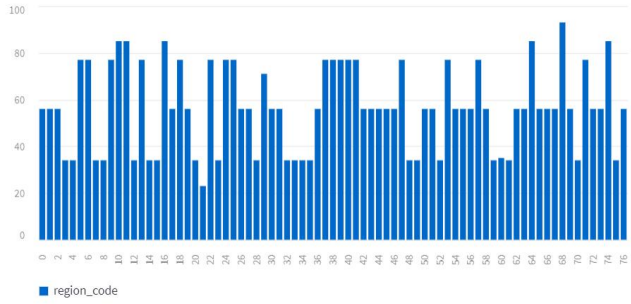


*Figure 5: Bar chart for region code*

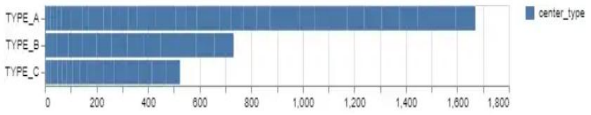


*Figure 6: Bar chart for the center type*

Figure 5 and 6 represent the bar charts from the second CSV file namely, fulfillment_center_info. The first bar chart shows the frequency of the region code whereas the second bar chart shows the horizontal bar chart with 3 center types as shown in the figure. The region code bar chart shows that the highest and lowest region codes are 93 and 23 at indexes 68 and 21 respectively. Similarly, the center type bar chart shows that type A and type C are the highest and lowest in frequency as shown above.

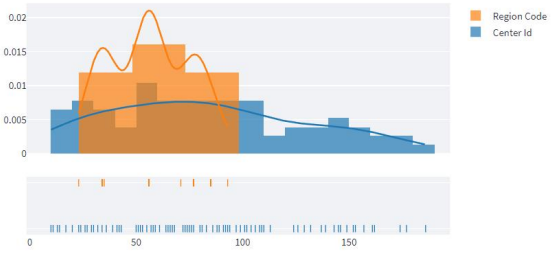


*Figure 7: Distribution region code and center id*

Figure 7 shows the distribution of region code and the center id in the fulfillment center info dataset. Here, the center id ranges from about 10 to 185 whereas the distribution of region code between the center id ranges from about 20 to 100.

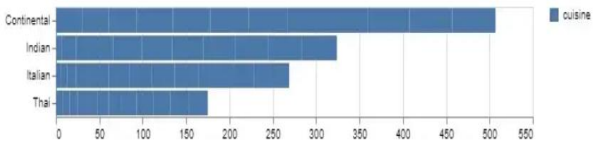


*Figure 8: Bar chart for cusine column*

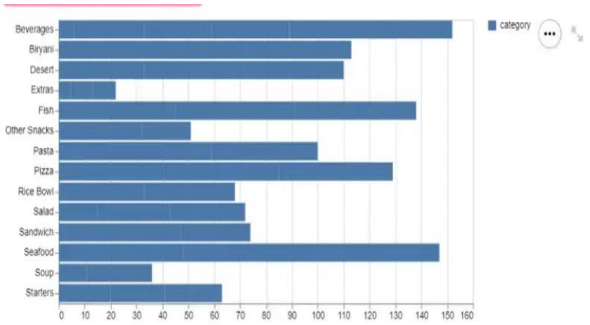


*Figure 9: Bar chart for meal categories*

Figure 8 and 9 shows the bar chart representation of the cuisine and meal categories from the third CSV file containing the meal information. From the first bar chart, we can see that there are four cuisines including Continental (50), Indian (40), Italian (41) and Thai (22). Similarly, the second bar chart shows the categories of different kinds of meals demaned by the customers. From the two charts we see that, the customers demand continental cuisines the most and the thai the least whereas they demand the seafood the most and the soup and extras meal the least.

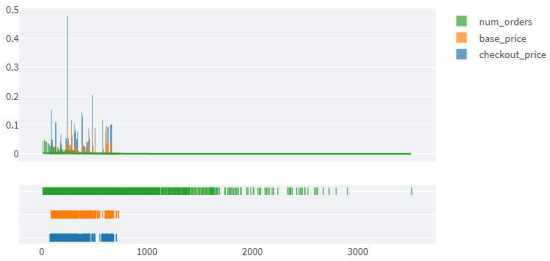


*Figure 10: Distribution for weekly demand data*

Figure 10 shows the distribution chart for the weekly demand data for 50000 rows of data. Unlike the charts plotted previously for the weekly demand data like bar, line and area charts, we can observe that the base and checkout price do not overlap and range quite differently expect in few points. As the more number of rows of data are observed we can see that the checkout prices are relatively higher as compared to base price in general. This shows that the rate of profitability is higher when we observe the higher number of weeks and data.

## VII. DISCUSSION

The following project is based on the datasets to forecast the food demand of the customers in the upcoming weeks based on pervious weekly demanded data in general. Here, our basic goal was to determine or rather predict what foods the customers are going to demand in the future so that the company can be prior ready in production and supply of that very product. Here, The replenishment of majority of raw materials is done on weekly basis and since the raw material is perishable, the procurement planning is of utmost importance. Secondly, staffing of the centers is also one area wherein accurate demand forecasts are really helpful. Given the following information, the task was to predict the demand for the next 10 weeks for the center-meal combinations in the datasets. Overall, we have observerd three CSV files including the weekly demand informantion, fulfillment center information and the meal information. From the different types of data visualization perfomed, we observed that the center types A and B have more food demand as compared to the center type C. Similarly, among the demand of cuisines from the restaurant or company, the customers have most prefered and hence demanded the continental cuisine. Besides, they also have a significant demand of Indian and Italian foods. However, the demand of Thai cuisines are not as much as compared to the rest of the 3 ones. Moving on to the categories of the meal, from the food demanded previously, in general most of the customers are likely to demand the foods like Biryani, fish, seafood, pizza along with the beverages. However, only few of the customers demanded soup, salad and the other snacks. This helps us predict that in general the customers prefer the non-vegeterian items, the beverages and the desert more than other items. We also observed the profitability of the restaurant within few weeks and within longer duration of weeks.

These quantitative analysis of food demand forecasting dataset included the key performance indicators in order to forecast the upcoming demands and hence produce and supply the food orders as per their demand in appropriate time with significant profit to the restaurant or respective company.

## VIII. CONCLUSION

Data mining can be utilized to find covered up, obscure, however valuable learning from huge, fuzzy, uproarious, fragmented, and irregular data. In this paper, we present our task of predicting the the food demand by the customers to ensure the growth of the restaurant or industry in well directed direction. By building a dashboard using streamlit we could understand, analyze and forecast the demanded food in an efficient manner.When you are working with Streamlit, you don't have to worry about your front-end knowledge. Streamlit framework will easily convert data scripts into a shareable web application with just a few lines of coding. From this project using the respective framework, we could understand more about the quantitative analysis and the exploratory data analysis. By implementing this

project, we could successfully conclude that data visualization provides a good, organized pictorial representation of the data which makes it easier to understand, observe, analyze. Not only in case of food demand forecast but ni any small or huge data they can be transformed into the visualized form so that they can be interpreted and analyzed to increase the profit of the company as well as improve the quality of any institution in the upcoming future by predicting their requirements.

## REFERENCES


[1] *Data Visualization Overview*. IEEE Xplore. (n.d.). Retrieved November 21, 2022, from https://ieeexplore.ieee.org/document/9633610

[2] "An Overview of Data Visualization." An Overview of Data Visualization | IEEE Conference Publication | IEEE Xplore, ieeexplore.ieee.org/document/9012031. Accessed 21 Nov. 2022.

[3] "What Is Data Visualization? Definition, Examples, and Learning Resources." Tableau, www.tableau.com/learn/articles/data-visualization. Accessed 21 Nov. 2022.

[4] C.Min et al., "Data, information and knowledge in visualization," IEEE Comput. Graph. Appl., vol. 29, no. 1, pp. 12-19, 2009.

[5] E.R. Tufte and G.M.Schmieg, "The Visual Display of Quantitative Information," Am. J. Phys., vol. 53, no. 11, pp. 1117-1118, 1985.

[6] "Data Visualization Dashboard | DashThis." Data Visualization Dashboard | DashThis, dashthis.com/data-visualization-dashboard. Accessed 21 Nov. 2022.

[7] "Dashboards for Data Visualization." Dashboards for Data Visualization, www.linkedin.com/pulse/dashboards-data-visualization-meghna-goyal. Accessed 21 Nov. 2022.

[8] "How to Build Data Visualization Dashboards | Sigma Computing." How to Build Data Visualization Dashboards | Sigma Computing, www.sigmacomputing.com/blog/how-to-build-data-visualization-dashboards. Accessed 21 Nov. 2022.

[9] "Streamlit • the Fastest Way to Build and Share Data Apps." Streamlit • the Fastest Way to Build and Share Data Apps, undefined. Accessed 21 Nov. 2022.

[10] "Data Visualization using Streamlit | Introduction to Streamlit." |https://towardsdatascience.com/data-visualization-using-streamlit-151f4c85c79a. Accessed 21 Nov. 2022.

[11] "Introduction to Streamlit | LatentView." LatentView Analytics, www.latentview.com/data-engineering-lp/introduction-to-streamlit. Accessed 21 Nov. 2022.

[12] "Identity Decisioning Platform for Financial Services | Alloy." Identity Decisioning Platform for Financial Services | Alloy, www.alloy.com. Accessed 21 Nov. 2022.

[13] https://digitory.com/. digitory.com. Accessed 21 Nov. 2022.

[14] "Global Demand Intelligence and Data Platform - PredictHQ." Global Demand Intelligence and Data Platform - PredictHQ, www.predicthq.com. Accessed 21 Nov. 2022.

[15] "Planungssicherheit Senkt Lebensmittelverschwendung." Planungssicherheit Senkt Lebensmittelverschwendung, www.delicious-data.com. Accessed 21 Nov. 2022.

[16] "Home Page." 2020_04_Perego, www.politesi.polimi.it. Accessed 21 Nov. 2022.

[17] "Datasets - Definition, Types, Properties and Examples." BYJUS, byjus.com/maths/data-sets. Accessed 21 Nov. 2022.