**Business Problem:**

* Predicting future demand for different items in a grocery store.
* [Corporación Favorita](http://www.corporacionfavorita.com/) is a large Ecuadorian-based grocery retailer which operates hundreds of supermarkets, with over 200,000 different products on their shelves.
* It is a timeseries analytics problem and there are various methodologies to solve these kinds of problems.
* The client, [Corporación Favorita](http://www.corporacionfavorita.com/), has come to an agreement for allowing the usage of their data for analyzing and coming up with a solution.
* Benefits of effectively predicting sales forecasts.
* **Efficient supply chain scheduling**: Forecasting the amount of sales and when they are likely to occur, would help better scheduling production, warehousing and shipping.
* **Better Labor Management**: Anticipating demand means knowing when to increase staff and other resources to keep operations running smoothly during peak periods.
* **Adequate Cash Flow:** Knowing the peaks and valleys of demand would help in better management of cash flow, ensuring there is enough money and bills are paid on time**.**
* **More Accurate Budgeting:** The more accurately you can forecast demand, including the timing of your sales, the more accurate you can be with budgeting.·        
  Reformulate problem statement as an analytics problem

**Analytics Problem definition:**

* This is a time-series, forecasting the amount of unit-sales for a given store and is given item number is being predicted for a period of 15 days. Different factors are taken into account while training our model. The oilprice, an indicator of global economy. If The given day was a holiday, if the products sold were on promotion, the region in which a particular store is present and so on and so forth.
* Stores are clustered based on similar types.
* As Ecuador is an oil-dependent country we’re assuming it's economical health is highly vulnerable to shocks in oil prices.
* Success would be in measuring how close our predicted values are to real values.
* Measures include Mean Forecast error, Mean absolute error, Root Mean squared error.

**Data:**

* Data over the amount of unit-sales, for a corresponding store number and item number has been collected over a period of 3 years, for all the stores.
* Data on Holidays has been collected, classified into their types, for different dates during which our store data is collected.
* Data of Oil prices over the same given period.
* Preprocessing
* Features coreced into their numeric or categorical types depending on characteristics. Dates to PosiXCT.
* The categorical features into factors and then into dummy variables using one hot coding and linear combinations are removed to overcome rank-deficit matrix/ Non-invertibility.
* The Features more than 15% of missing values are removed and features less than 15% missing, have been imputed using “cart” method from the mice package.
* The Highly correlated features have been removed.
* The Variable Importance function - To find the importance features used in the model.
* Document and report findings
* Saturdays are the most popular day for shopping. Sundays are the second most popular day, with rest of the weekdays being pretty similar
* April-July(Summer) have increased sales as well as December due to the Christmas effect. Surprisingly, September to November seem to have much lower sales.
* Weather pattern does not seem to be correlated to the trends seen in the heatmap.
* The first few days of the month seem to have increased sales which could be explained by wages in the public sector being paid every two weeks on the 15th and last day of the month.

**Methodology Selection**

* Identify a few problem solving approaches (methods) and use one or two
* Mostly used methods include single artificial neural network (ANN) models with one or two traditional methods such as the exponential smoothing, moving average (MA), ARIMA, seasonal ARIMA (SARIMA), and generalized autoregressive conditional heteroscedastic (GARCH) models.
* Test and select an approach you believe might work
* The methods we’ve employed our model are ARIMA, XGBoost, ETS (Exponential Smoothing State Space Model)
* We applied the above given models for Store Number 44 and Item Number 329362.
* ETS: Exponential Smoothing State Space Model

**Model Building:**

* Models used for our analysis: ARIMA, ETS and PROPHET
* Run and evaluate the model
* Prophet is a relatively new method of forecasting which is developed by facebook’s core data science team. Prophet fits non linear trends with yearly, weekly, and daily seasonality.
* prediction of daily unit sales for items at store level, for Store 40 and Item 314393.
* Train Set: 1st Aug 2016 to 31st July 2017, Test Set: 1st Aug 2017 to 15 Aug 2017
* key performance measures: RMSE. MAX RMSE = ARIMA, MIN RMSE = PROPHET.

**FUNCTIONALITY:**

* Shiny app: user interactive, can dynamically change the output. Multiple filters(user inputs)  were placed in the and side panel and it directly alters the final forecast graph based on the applied filters.
* We have created Various types of inputs in the user interface like Drop box , Selecting the dates by creating calendar interface.
* we have created multipurpose input panel , where we can either input the value or select from the drop down .
* Multiple subsetting of the dataset is done using nested reactive functions .
* A final graph was plotted and sent to UI from Server using Output$Plot
* Text is pasted in the  final output using Render function .
* useful packages like shiny themes, shinyDashboard, shinyWidgets while building the app. ShinyDashboard makes it easy in R to create multiple filters in different pages filtering an common dashboard

**GUI Design and Quality**

* additional features: Drop down features and Message, Menu’s Shiny Widgets play a major role in building dynamic UI that reacts to user input.
* conditionalPanel creates a panel that shows and hides its contents depending on the value of a JavaScript expression. Even if you don’t know any JavaScript, simple comparison or equality operations are extremely easy to do, as they look a lot like R (and many other programming languages).
* There are two main parts in RSS - 1)User Interface (UI.R), 2)Server Function (Server.R)
* In the User Interface part, we have made use of the shiny widgets to align the output interface . The Key functions in the shiny widgets like setting the side panel ,Main Frame made our coding part a  good experience.
* In the server part, in order to exercise multiple inputs on the output plot,  we have to use nested reactive functions.
* We created conditional statements to impart successive filtering .We have also used   loops to dynamically input values into an empty data frame and then to successively plot the graph.