Project Report

Market Revenue for Beverage Company : Orange Power

Problem Statement - Client Orange-Power wants to increase its footprint in the business of beverages. What should be the strategy of the company to make them successful by increasing market share for sales and revenue by at least 10%.

Data - Data set has been taken from Kaggle from the below mentioned link
Link - https://www.kaggle.com/veeralakrishna/predict-demand?select=train.csv
It consists of sales data for multiple beverage companies based in Greece.
Dataset consists of 11 features presenting sales data collected from different store for the all the beverage companies.

Dataset Overview :-

Total Rows - 6480 rows
Columns - 11
Sales data has been collected for each shop at the end of the month expanding over 5 years.

Data Types :

Data types for the given dataset were correctly populated.

Feature	Datatype
date	datetime64[ns]
city	object
lat	float64
long	float64
pop	float64
shop	object
brand	object

container	object
capacity	object
price	float64
quantity	float64

No action was required on this front.

Data Cleaning :

Following was initial dataset info snippet :

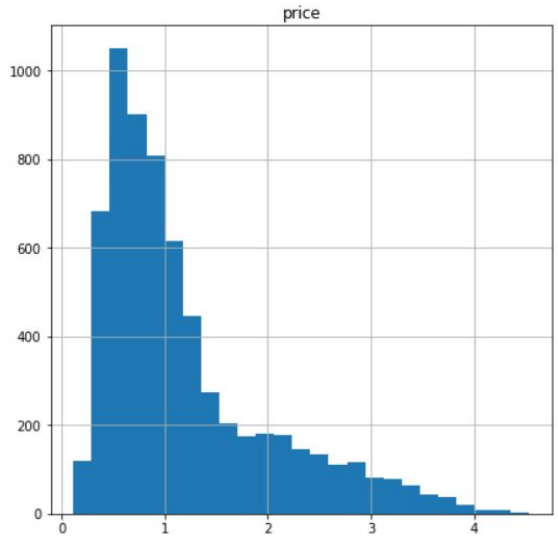
```
Float64Index: 7560 entries, 0.0 to nan
Data columns (total 11 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   date       6480 non-null   datetime64[ns]
 1   city       6480 non-null   object
 2   lat        6429 non-null   float64
 3   long       6434 non-null   float64
 4   pop        6480 non-null   float64
 5   shop       6480 non-null   object
 6   brand      6480 non-null   object
 7   container  6464 non-null   object
 8   capacity   6465 non-null   object
 9   price      6480 non-null   float64
 10  quantity   6480 non-null   float64
```

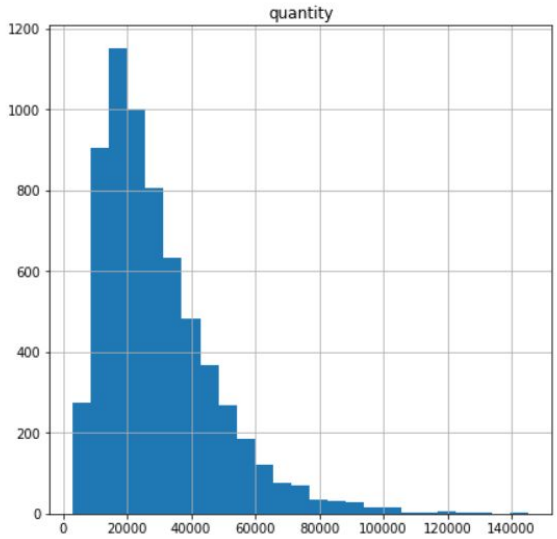
There were taken care of as following:
Lat - The column has been updated based on the relevant cities
Long - The column has been updated based on the relevant cities
Capacity - The column has been updated based on the type of container they belong to eg . can has 330 ml capacity while glass has 1000 ml and plastic has 1500 ml capacity.
Container - The missing values were updated based on the corresponding capacities.

Exploratory Data Analysis:

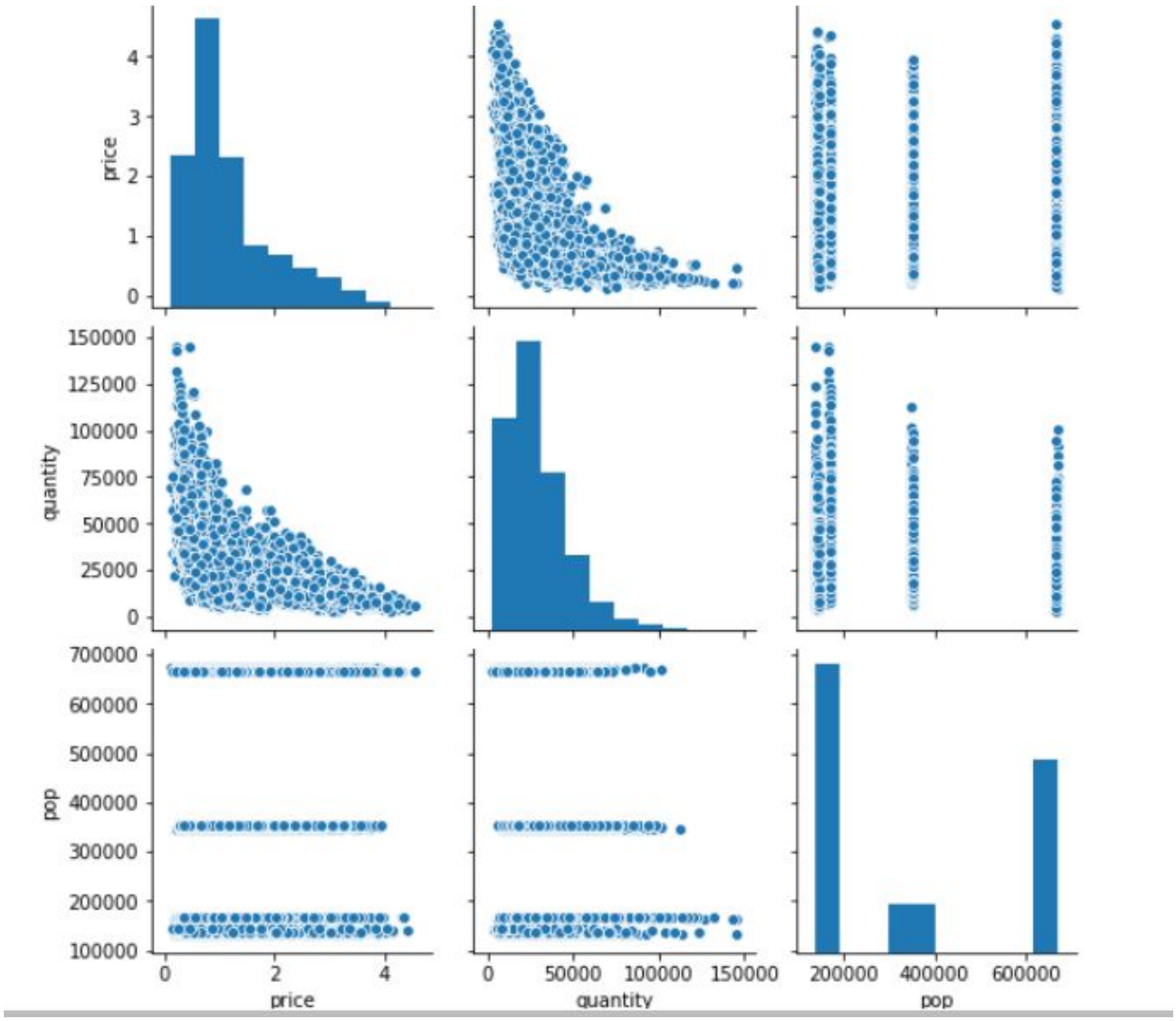
As sale of beverages depends upon the month of sale so we add a new column **‘month’** signifying the month in which the given record has been created.

Histogram of the data to have a view regarding the frequency spread for ‘Quantity’ and ‘ price :
We can see a clear right tail in both the variables.
Prices for the beverages mostly ranges below 1.5
Quantity of sales are also concentrated below 5000 and occasionally going above 7000 which might be for few months of higher sales (seasonality effect).



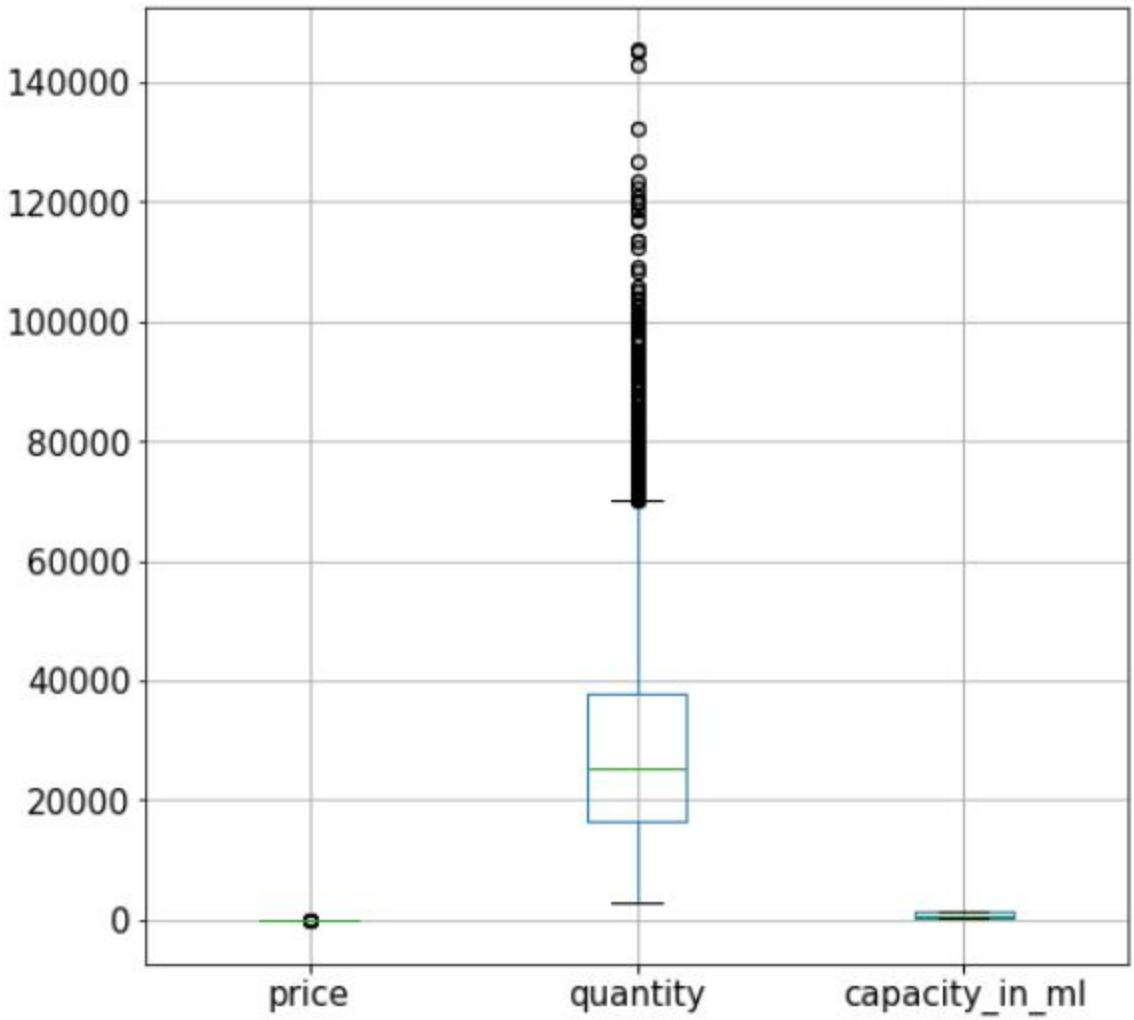


Pairplot for the relation B/w features.[For Numerical variable] :



On analysing the above graph we can see a clear negative correlation between **‘price’** and **‘quantity’** which gives an idea about a price sensitive market when increase in prices leads to decrease in the quantity of sales.

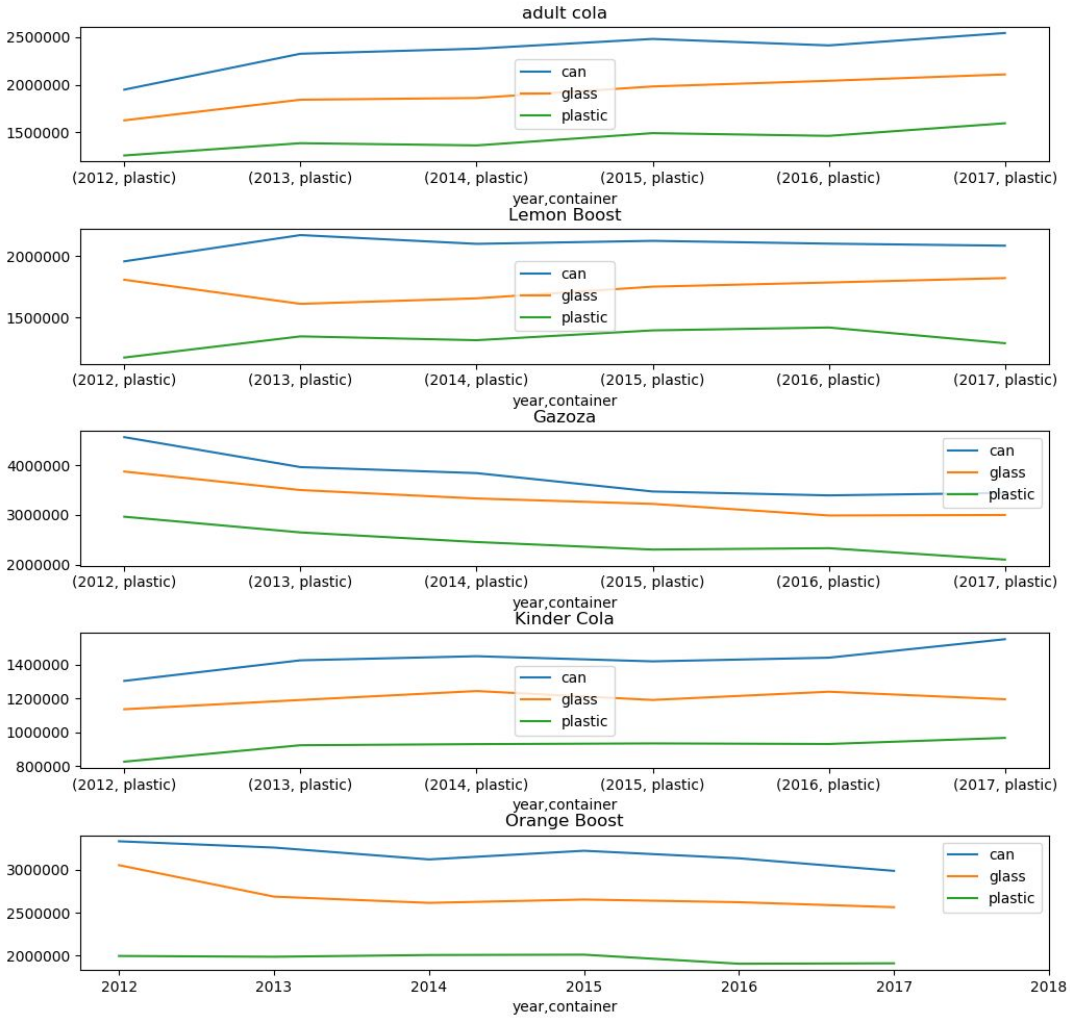
Box Plot for identification of Outliers:



In the above presented graph we can see a clear outliers for quantities but these won't be consider as actual outlier because of the nature of the variable i.e. 'quantity' which signifies the sale of a given variant for a given company. This shows market dominance in terms of sales for a particular brand .

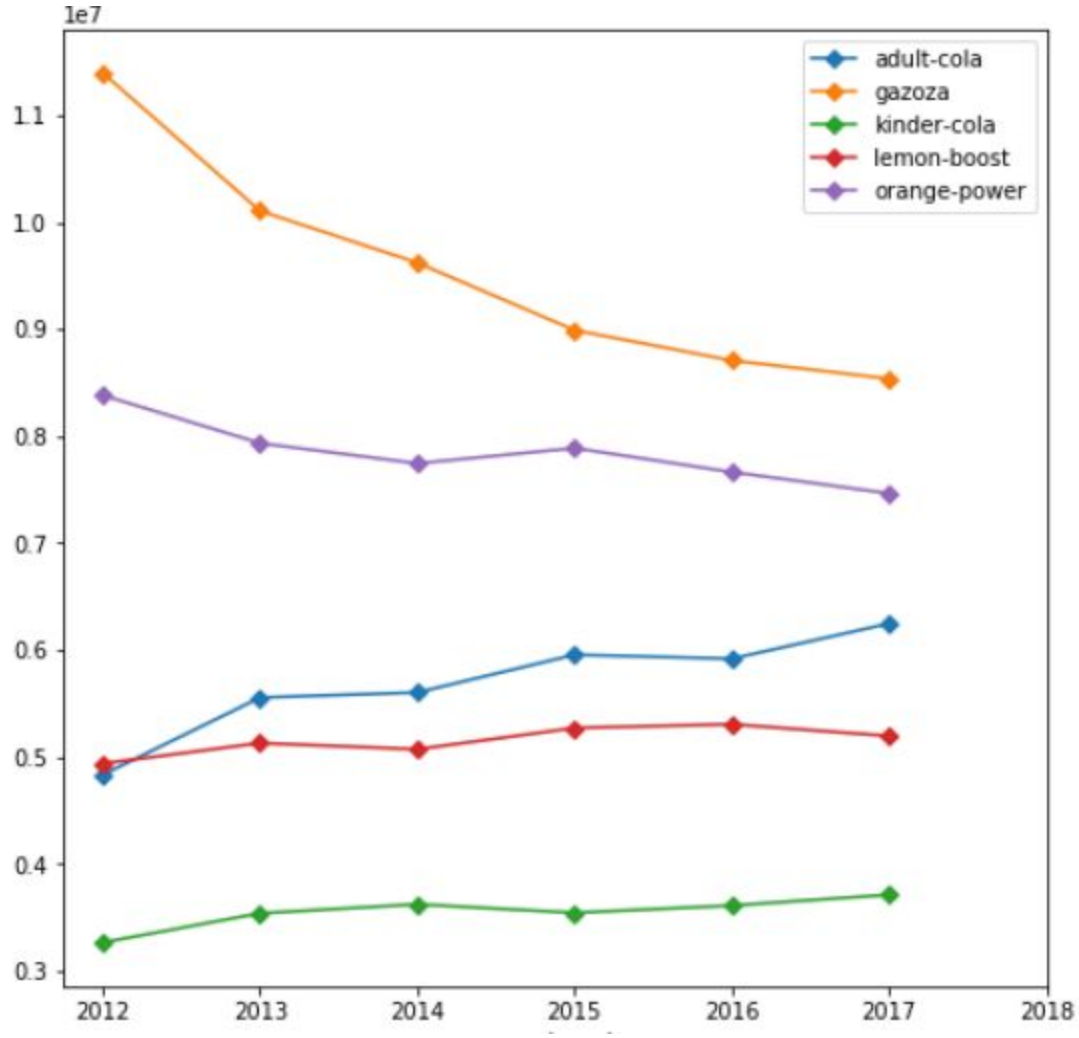
Trend Analysis :

As we have inspected the data initially and we can see that data provided is in the of time-series.
We are analysing sales trend for different beverages over a period of time :



The above diagram shows how sales of different variants of given beverages changed over time.
We can infer that sales for **orange boost** have had a steady decline over the period of time.
Largest decline in the sales is for **gazoza** showing a larger negative slope.
Adult cola is the beverage which has shown a promising upward trend in terms of sales.

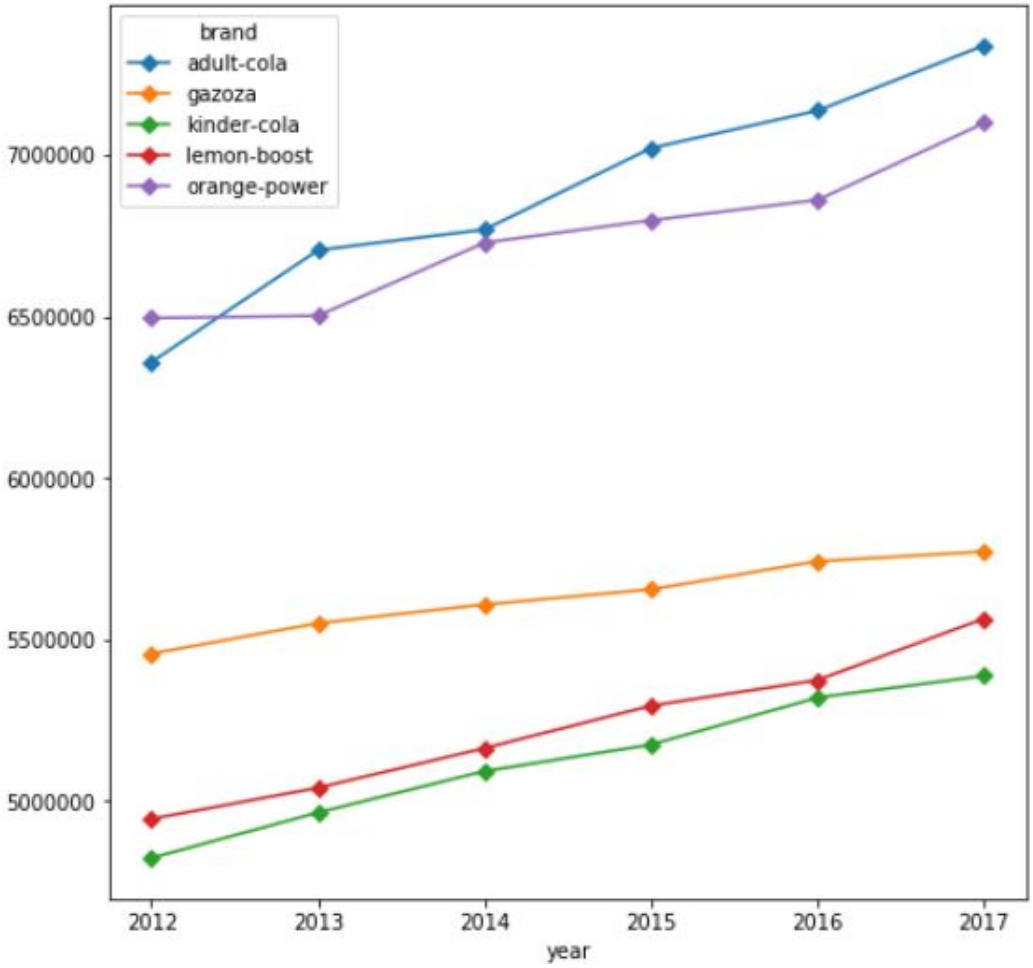
We further analysed total sales for the beverages over the period time.



Here we can again infer that over the period of time our client i.e. **'Orange-Power'** has shown decline in sales.
Other important inferences are : Gazoza has the largest numbers in sales despite the constant decline over the period.
Adult cola shows promising increment in sales closing its gap with its competitors at a rapid pace.

In earlier visualizations , we have been focusing on the sales component , as companies try to increase their total Revenue which is governed by two factors Sales and Price.

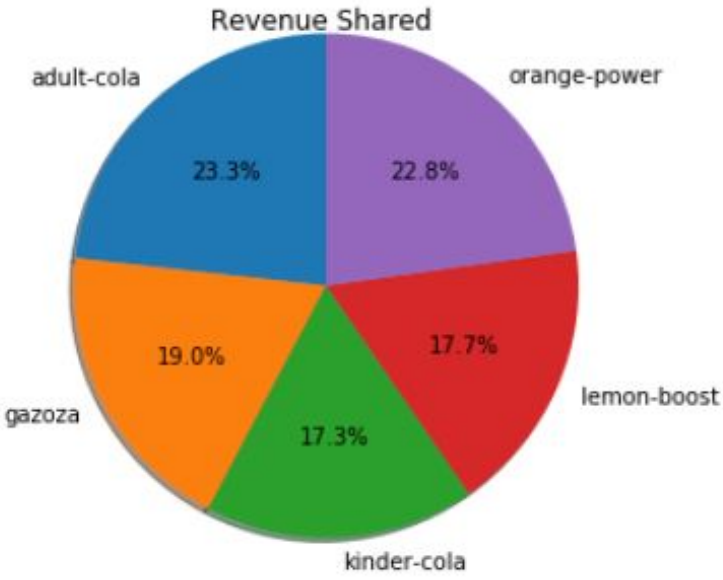
Below Fig. shows the change in revenue of different companies over the period of time

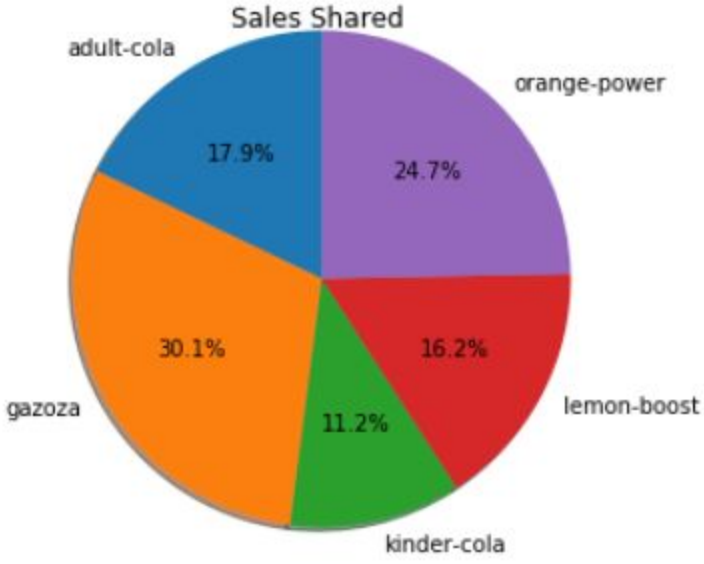


In contrast to the sales trend depicted earlier , here we see increment in the revenue of companies with Adult-cola leading the market followed by Orange-Power .

Gazoza which has the highest number of sales is lagging in the revenue front.

From above visualization we have seen individual beverage companies' sales and revenue trends.
Let us see how well these companies fair in the market by visualization their share in the market .





The above pie chart shows the market share for all the companies in terms of Sales and Revenue.
Adult Cola leads the revenue with 23.3% share trailed by 'Orange-Power' with 22.8%.
Adult Cola leads the revenue with 30.0% share trailed by 'Orange-Power' with 24.7%.

Our Project aim is to elevate the Orange Power market by increasing its presence in the market w.r.t. To sales as well as revenue.

Model Training and Prediction:

We have to define dummy variables for categorical Items to consider their impact. Two datasets were created one considering the ‘years’ impact on the prediction while other one considering the seasonality to counter the changes in period of time.

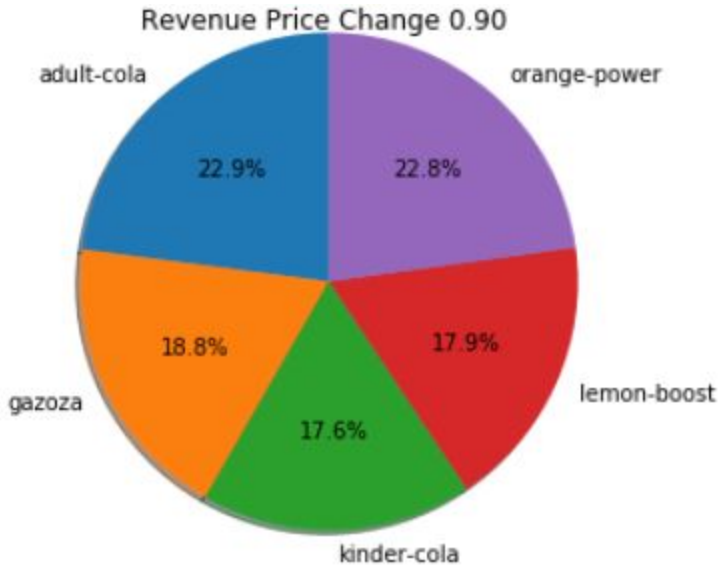
We have created models based on Linear regression , Random Forest and XG_Boost . Result were as below :

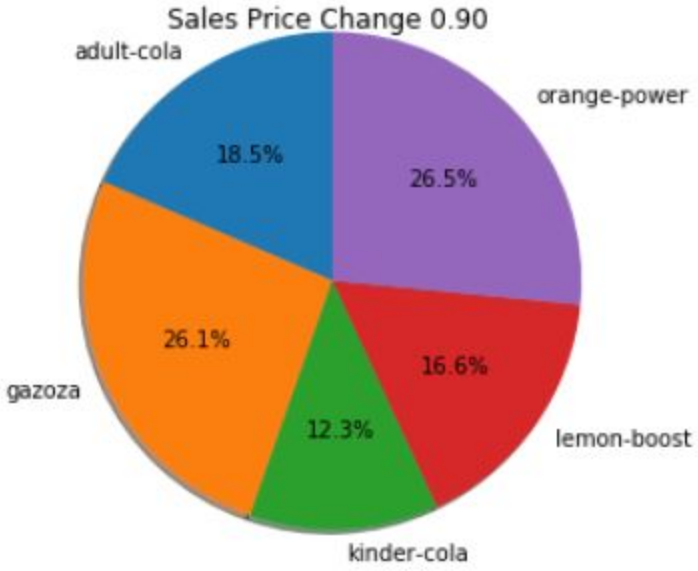
Results

Model	Data set	R-Squared	Mean Absolute Error
Linear regression	With year	0.731	6629
Linear regression	w/o year	0.719	6721
Random Forest	With year	0.915	3421
Random Forest	**w/o year**	**0.926**	**3144**
XG Boost	With year	0.887	4118
XG Boost	w/o	0.883	4154

Random forest showing better results as compared to other models.

Prediction: Using the above selected model and making variance in the price provided initially by the client.
If the price is decreased by 10% of the proposed value provided by ‘Orange-Power” , the company can increase its share in Sales by more than 15% while closing the gap in terms of revenue share to 0.1% from initial 0.5%.





Conclusion :

As per our analysis we can suggest that the price provided by the client for the upcoming year would maintain the status quo in terms of market share Revenue as well as the Sales.
We have done predictive analysis with different pricing levels based on the input provided and below are the findings:

* With decrease in price will lead to increase in sales.
* Revenue being a function of Sales and Price will be showing initial increase with decreasing price due to hike in sale quantity . Revenue will get optimized at a price difference of 0.90.

* Market is price sensitive as increasing the prices shows decrease in both the Revenue as well as sales of beverages.
* **Orange-Power** is seen to be most optimized in business terms as it is leading in both sales and revenue striking a good balance as compared to its competitors i.e. Adult-cola which is leading in Revenue share but lags drastically in terms of its sales on the other hand Gazoza is leader in sales but lags on Revenue front.

Scope Of Improvement :

* Model can be improved by considering the changes in the inflation rate which would help me capture changes with years passing by.
* Population increment is another factor which can be weighed in to further improve the model prediction.
* We can also have an apply strategy to give seasonal discount as the sale of beverages shows a seasonal trend rather than applying flat discount throughout the year.