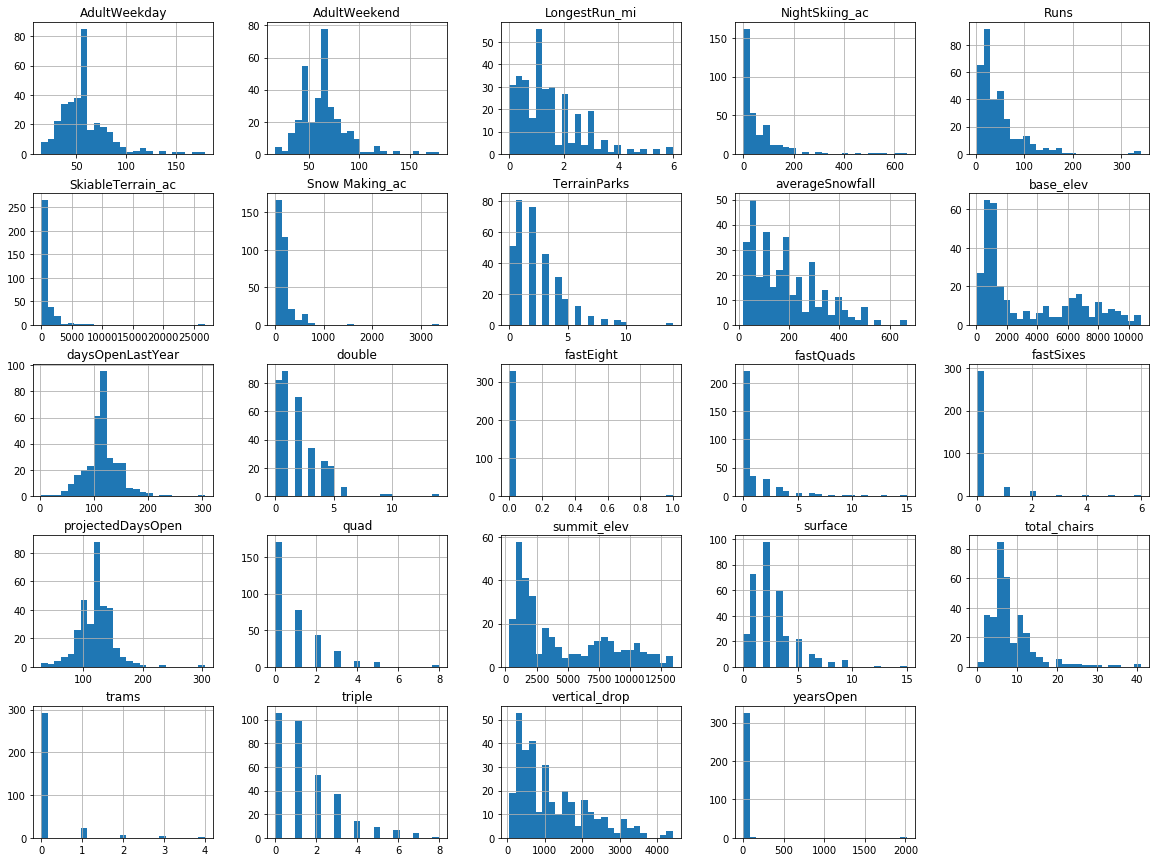
***BIG MOUNTAIN******RESORT PROJECT REPORT***

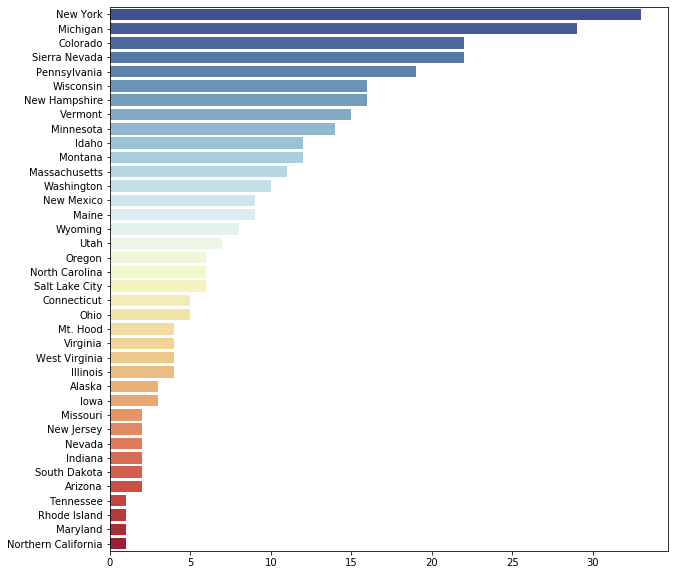
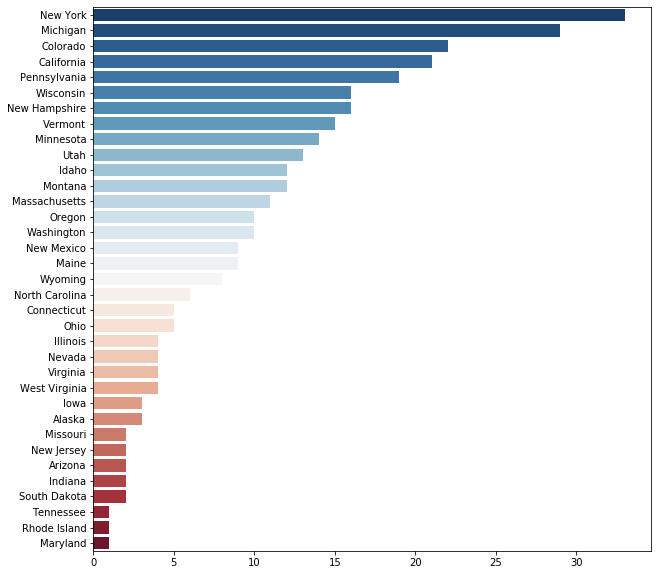
***PROBLEM STATEMENT :*** What changes must be implemented by Big Mountain Resort to improve its business profits to compensate for the increased operational costs $15,40,000( new chair lift along with existing lifts) and to maintain a current profit margin of 9.2% in this annual revenue?

***DATA OVERVIEW*** ***:***

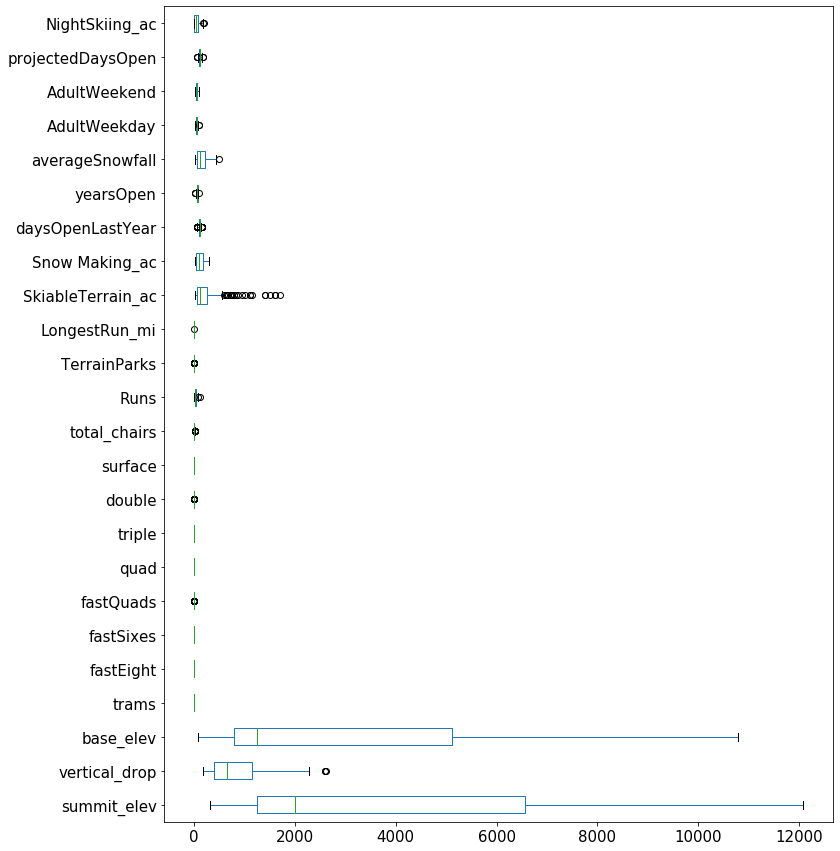
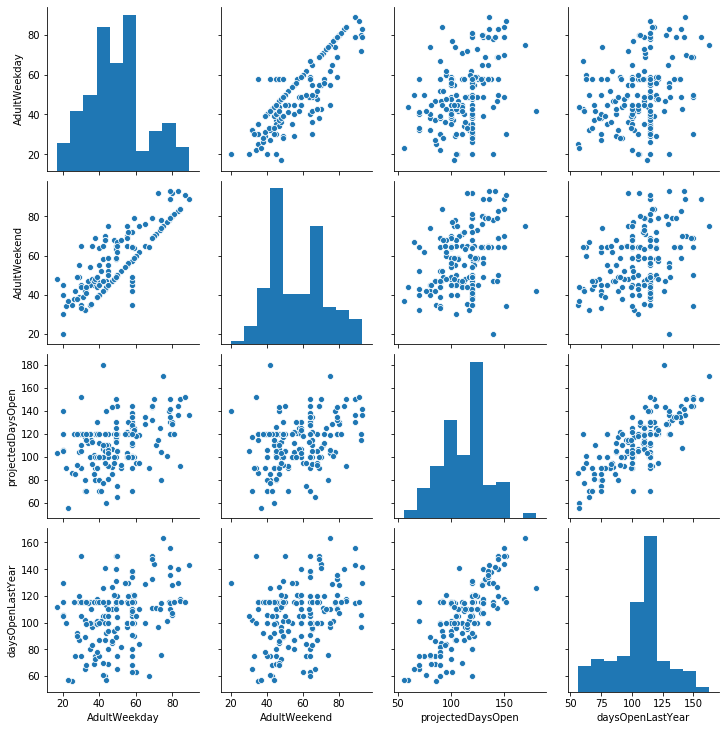
* As a part of data collection, data was collected from “updated\_ski\_data.csv”, it contains 27 columns and 330 rows which represents features of skiing resorts and names of the regions respectively.
* Data organized by file structuring in to ‘data’,’models’,’figures’ by using ‘mkdir’ command.
* Column names,Data types,Description of the columns,Count of unique values and range of values reviewed in Data definition part.
* Data cleaning process initiated by analysing null values, missing data and duplicate values. In this data set i found ‘Fasteight’ lift column as null column which was dropped using ‘isna’ command and missing data in ‘NightSkiing\_ac’ column filled with ‘0’.

***EXPLORATORY DATA ANALYSIS*** ***:***

* Builded the dataset profile plots and tables to analyse the outliers and anomalies, found State and region column as anomalies with help of level plots and with box plot and inter quartile range found SkiiableTerrain\_ac as outlier.

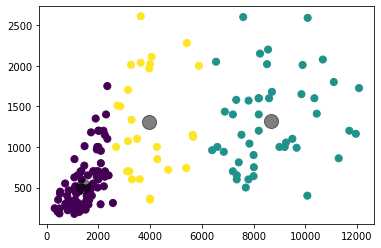
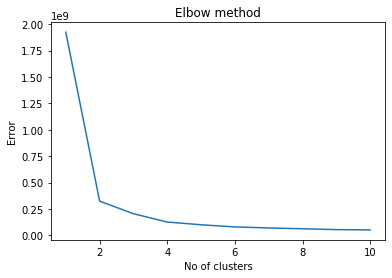


***Histogram plots of all columns State Level Plot Region Level Plot***

******

***Box Plot for Outliers Data Relationship Scatter Plot***

* From pearson correlation coefficient heat map, the column with threshold > 0.95 was ‘base\_elev’ is dropped
* After removing the outliers and Region column dropped as its values are same as column, the shape of the dataset from (330,27) to (176,25).
* Elbow plot was plotted inorder to create cluster to find the patterns among the dataset.



***Elbow Method***  ***Cluster for Dataset***

***PRE-PROCESSING AND DATA TRAINING :***

* Created dummy variables or indicator features for categorical variables
* Standardized the magnitude of numeric features
* Splited data into testing and training datasets.
* Apply scaler to the testing set