JAYPEE INSTITUTE OF INFORMATION TECHNOLOGY



**MID MINOR REPORT**

**DEAL ESTIMATION**

|  |  |  |
| --- | --- | --- |
| **S.no** | **Name** | **Enrolment No.** |
| 1 | Abhijit Pratap Singh | 15103286 |
| 2 | Ashi Sinha | 15103291 |
| 3 | Somika Rastogi | 15103309 |
| 4 | Yash Kumar Singhal | 15103315 |

Submitted To:-

Dr. Adwitya Sinha

(JAYPEE INSTITUTE OF INFORMATION TECHNOLOGY)

**PROBLEM STATEMENT AND PROPOSED SOLUTION**

In the world where there are fluctuations in prices of the products on daily

Basis, there is a growing need for predicting the optimal market price of the

product and thus the project would prove beneficial for seller. The seller

might want to know the best price he could sell his product and earn a profit

given the descriptors related to the product (Condition, Item,

Brand, Advertising Tag) and the current market want of the product

(Seasonal Trend) . The price would also be affected by how the seller would

advertise its product, recommending him whether he should stress upon

the advertisement to earn the profit more . Therefore the application is

capable to consider these changes and provide the most optimal price.

* Dataset used for the project has been manipulated and visualized to work on it. Columns“crawl\_timestamp","Col1","retail\_price","discounted\_price","is\_FK\_Advantage\_product","brand" are used for implementing feature of predicting accurate price of the product for the seller depending on the values in these columns.
* Model has been trained for predicting the price using Polynomial Regression.
* For better understanding of how the model estimates the price , statistical analysis of the graphs were made built on various parameters:

1. Demand Vs Category
2. Category Vs Retail price
3. Retail price Vs Demand
4. Months Vs Price depending upon category
5. Demand Vs Months depending upon category

* For taking input from the user and based on that visualization of the graphs,price prediction is implemented using Shiny Framework of Rstudio.

**TECHNOLOGY USED**

|  |  |  |  |
| --- | --- | --- | --- |
| S.no | Platforms and Concept Used | Description | Usage in our project |
| 1 | Rstudio | RStudio is an integrated development environment (IDE) for R. It includes a console, syntax-highlighting editor that supports direct code execution, as well as tools for plotting, history, debugging and workspace management. | dplyr library used for data manipulation.  Functions like groupby(),separate(),filter()  have been used to clean the data.  Filter() is used for taking all those column values that have the maximum count in the dataset.  Separate() is used to split the category of the product into main category and sub-category.  Groupby() used to group the category of the data on basis of timestamp and them taking its mean to get the seasonal demand and price for that category.  Stringr library- functions like str\_replace\_all(), str\_extract\_all(),paste(),sapply(),  str(),substring() helped us in tidying up messy string into possible usable formats  ggplot2 library- functions like result.geom\_boxplot(), geom\_point(), geom\_bar(),geom.\_line(). Helped us in plotting the required graphs |
| 2 | Shiny | Shiny is an R package that makes it easy to build interactive web apps straight from R. | UI of the shiny app is set using the functions fluidPage(), which includes many of the widgets like sidebarLayout, titlePanel, textInput, helpText, selectInput, sliderInput, checkboxInput, actionButton.  The output for the demand graph with category and retail price is taken as output in value8,the price prediction is taken in value5 and the seasonal plotting of the categories output is taken in value6 which is then rendered and displayed  The Input from all of the widgets is taken and is accesed using the input$.  The server Function heled us to integrate The UI with the logic |
| 3 | Machine Learning | **Machine learning** is a field of [computer science](https://en.wikipedia.org/wiki/Computer_science) that gives [computer systems](https://en.wikipedia.org/wiki/Computer_systems) the ability to "learn" (i.e., progressively improve performance on a specific task) with [data](https://en.wikipedia.org/wiki/Data), without being explicitly programmed. | Since our Problem of concern was to display the continuous value , the non linear regression was probably the most optimal choice for the given model.  The high degrees of freedom allowed us to use polynomial regression to predict the expected prices without worrying about overfitting  Glm() function allowed us to seamless make the required model |

**RESULTS**

**The model was successfully trained using non linear Polynomial Regression and is now capable of predicting the expected retail cost given brand, maximum concession, and wether the product is ecommerce verified or not**

Stastical Summary of Prices on Category

|  |
| --- |
| **d1.png** |
| Fig 3.1 The graph of retail\_price Vs Category analysis tells us that retail price of category Furniture is highest whereas watches has minimum retail price variation. |

|  |
| --- |
| **d1.png** |
| Fig 3.3 This graph shows scatterplot between Demand and Retail\_price showing that as the demand increases for a particular category, the retail\_price increases but demand is clustered when retail\_price is in the range of 1000-2000. |

|  |
| --- |
| d1.png |
| Fig 3.4 This graph of Count Vs Is Advantage tells us that maximum of the products in our dataset is not verified by Flipkart commercial site and so the benefits of being verified are enjoyed only on few products whose count are in the range 0-500. Fig 3.4 This graph of Count Vs Is Advantage tells us that maximum of the products in our dataset is not verified by Flipkart commercial site and so the benefits of being verified are enjoyed only on few products whose count are in the range 0-500. |

|  |
| --- |
| d1.png |
| Fig 3.4 This graph of YY$Price Vs YY$Months helps us to analyze seasonal changes in price of the category "Bags".The hike in price is observed in the month of January and minimum price in February. |

|  |
| --- |
| d3.png |
| Fig 3.5 This graph of YY$Price Vs YY$Months helps us to analyze seasonal changes in price of the category "jewellery".The hike in price is observed in the month of January and minimum price in February. |

|  |
| --- |
| d2.png |
| Fig 3.6 This graph of YY$Demand Vs YY$Months helps us to analyze seasonal changes in demand of the products belonging to the category "jewellery" which shows its hike in month of June and lowest demand in January. This is well consisted with Fig 3.5 where price is highest in the month of January |

**FUTURE WORK:**

* Using Natural Language Processing, the application would analyze

the description tag that the seller uses for his product and whether

there will be price hike or fall if he changes his decription tag . In that

way the seller can work upon the best description tag for his product

so as to increase the market value.

* Connection of the current work with android application, using the plumbr library as the REST/API and Shiny