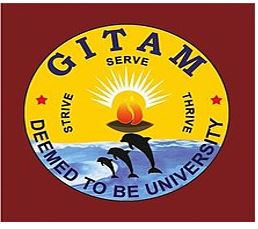


PURCHASE PREDICTION(BOOK)



BY,

SOHIT TALASU

GURU VARDHAN REDDY

SRISHTI SINGH

ESWAR CHAITANYA M

TEJASWINI YADAV N

**CERTIFICATE**

This is to certify that **SOHIT TALASU, GURUVARDHAN REDDY G,** **SRISHTI SINGH, ESWAR CHAITANYA M, TEJASWINI YADAV N** enrolled in the Summer Internship Program 2019 of the Smart Bridge Private Limited, Hyderabad has successfully completed a summer project entitled ‘**Purchase Prediction (Book)**’ during the time period from 13th May 2019 to 1st June, 2019. under the guidance of **Mr. Venkata Rao**, Smart Bridge Private Limited, Hyderabad.

Date:

Signature of Supervisor

**ACKNOWLEDGMENTS**

I hereby thank **Mr. Venkata Rao** who has given me the great opportunity to work on this project. He has been a great source of inspiration and his timely support and guidance has helped in the successful completion of the same. Then I would like to thank **Ms. Moulika,** her enthusiastic approach and dedication which has indeed been a great source of inspiration and support to me. She deserves a real acknowledgment at this juncture. Also, I express my sincere thanks to **Mr. Rahul** who guided me throughout the project.

I thank the department of Smart Bridge, Hyderabad for providing me with necessary infrastructure to do my work.

**Table of Contents**:

1) Title of the project

1.1) Introduction

1.2) Objectives of research

1.3) Problem statement

1.4) Industry profile

2) Review of literature

3) Data collections

4) Methodology

4.1) Exploratory data analysis

4.1.1) Box plots

4.1.2) Plots for sample input and output

4.1.3) Seaborn

4.1.4) Dist plots

4.1.5) Heatmap

4.1.6) Stats model

4.1.7) Scatter graphs

5) Findings and suggestions

6) Conclusion

7) Biography and reference

**1.1 INTRODUCTION**

Identifying what customer needs and how likely they are going to purchase the same or relative item again is a very crucial thing to consider while doing business. This aspect becomes really important when the competition and demands are always changing.

We have to make sure that no customer leaves our company as well as new customers begin to buy from us. It has also been observed that keeping old customers generates more profit than attracting new ones.

Purchase decision process describes the sequence of actions performed by a customer when deciding to purchase a particular product or service. It can also be described as a process of problem solving, where a consumer satisfies his needs after thoughtful consideration. The outcome of a purchase decision process is a decision whether a customer will buy a given product or service or not.

The purpose of this thesis is to study, implement and analyse various Machine Learning models and then do an analysis of the sample / raw data to obtain a meaningful interpretation. Some of the Machine Learning algorithms I have used, are Logistic Regression algorithm, Linear, multi linear algorithm.

* 1. **OBJECTIVES OF RESEARCH**

The main objective of this project is to develop a platform which will be simple and easy to use, as here one must provide the details of the different types of books, relatively purchased books and based on the features extracted the algorithm. The given dataset is a weak dataset as it is raw and unstructured. So, it cannot give a unique solution. As here, ridge algorithm does the task hence a well-trained model is less bound to make errors in predict whether the book florence is being purchased by max customers or not.

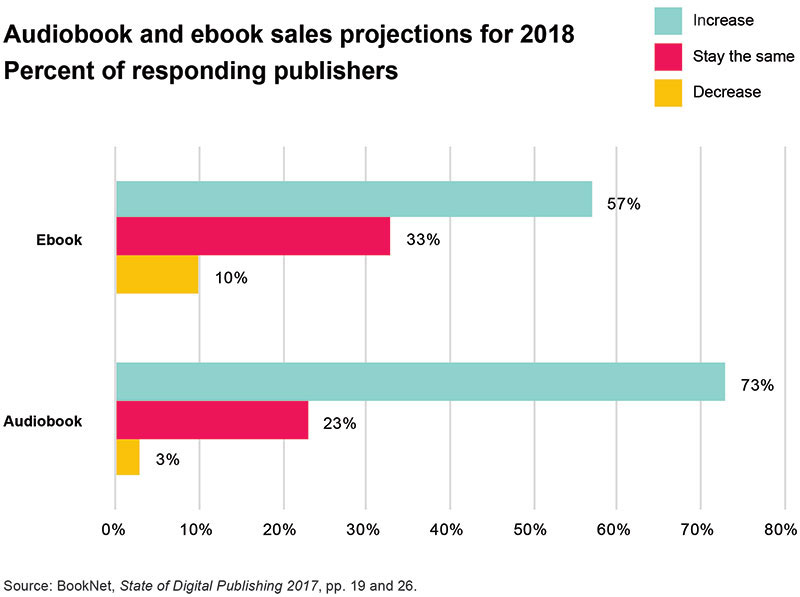
* 1. **Problem Statement**

**MAIN:** To predict whether the book florence is being purchased by max customers or not

OTHER EG:

* TO PREDICT HOW MANY CHILD BOOKS ARE PRESENT IN TOTAL NUMBER OF BOOKS
* TO PREDICT HOW MANY YOUTH BOOKS ARE PRESENT IN TOTAL NUMBER OF BOOKS
* TO PREDICT HOW MANY COOK BOOKS ARE PRESENT IN TOTAL NUMBER OF BOOKS
* TO PREDICT HOW MANY DOITY BOOKS ARE PRESENT IN TOTAL NUMBER OF BOOKS
* TO PREDICT HOW MANY BOOKS ARE PURCHASED FIRST AMONG TOTAL NUMBER OF BOOKS

* 1. **INDUSTRY PROFILE**

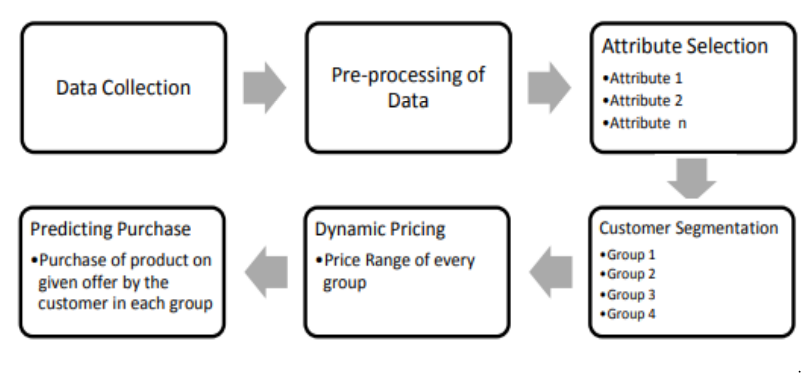
****

**2. REVIEW OF LITERATURE**

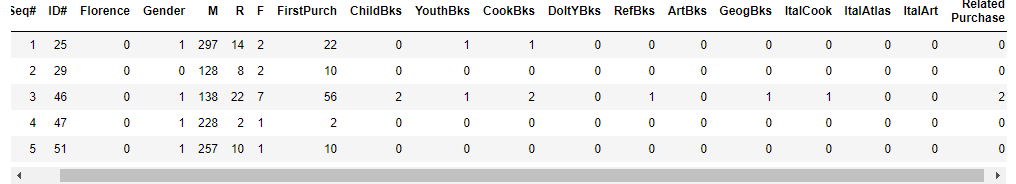
After searching and evaluating, the first outcome will be the decision to purchase or not to purchase the alternative evaluated as most desirable. If the decision is to buy, a series of related decisions must be made regarding the features, where and when to make the actual transaction, how to take delivery or possession, the mode of payment and other issues. So, the decision to make a purchase is really the beginning of an entirely new series of decisions that may be as time consuming and difficult as the initial one. Selecting a source from which to make a purchase is one of the buying decisions (Stanton, Etzel and Walker, 1994).

A consumer’s decision to modify, postpone or avoid a purchase decision is heavily influenced by perceived risk. The amount of perceived risk varies with the extent of money at stake, the amount of attribute uncertainty and the amount of consumer’s self-confidence. Marketers must understand the factors that provoke a feeling of risk in the consumer and provide information and support to reduce the perceived risk (Kotler, 2003).Peterson, Balasubramanian and Bronnenberg (1997) forecast that early in the twenty first century consumers will be purchasing food and other basic household needs via in-home television computer systems. The shopper will make choices after viewing brands and prices on the screen. So, the purchasing process itself may change dramatically in the coming decades. Lilien, Kotler and Moorthy (1999) reveal that a consumer’s purchase intention is influenced by changes in anticipated situational factors. The consumer forms a purchase intention on the basis of such factors as expected family income, the expected total cost of the product and the expected benefits of the product. Furthermore, when the consumer is about to act, unanticipated situational factors may intervene him from so (such as lack of availability of a preferred product). Hence, preferences and purchase intentions are not completely reliable predictors of actual buying behaviour : while they guidepurchase behavior, they fail to include a number of additional factors they may intervene. Demographics are uncontrollable variables in the external environment. The basis for any market is people. Hence, studying the population in terms of its demographic structure is very significant for marketing manager

**3. DATA COLLECTION**



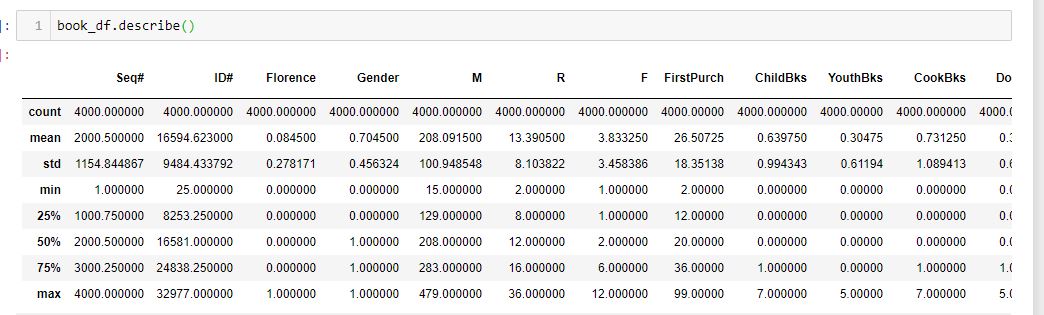
The current data set provides us with a data about the purchase of books that is provided by Mentors as book\_prediction.csv in which it contains 4000 rows and 19 columns.



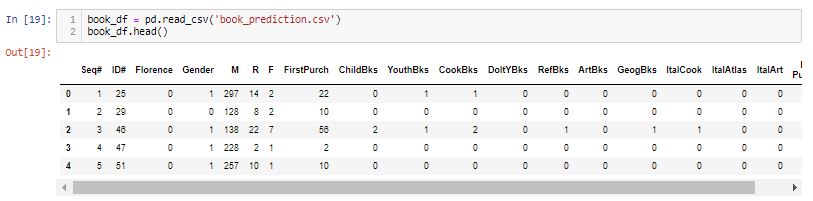
**4.METHODOLOGY**

**4.1) Exploratory Data Analysis:**

**Describe ()** is used to view some basic statistical details like percentile, mean, std etc. of a data frame or a series of numeric values.

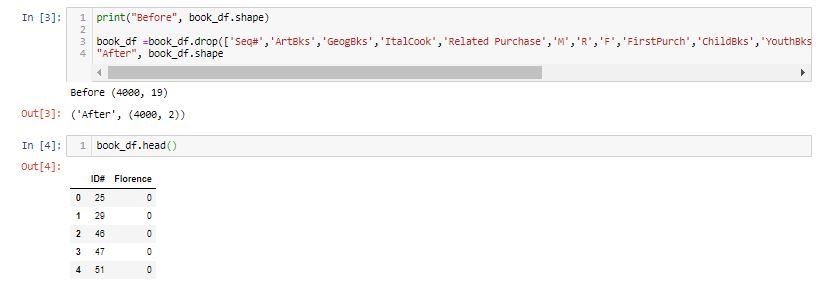
****

**head ()** is used to view first five values of the datset.

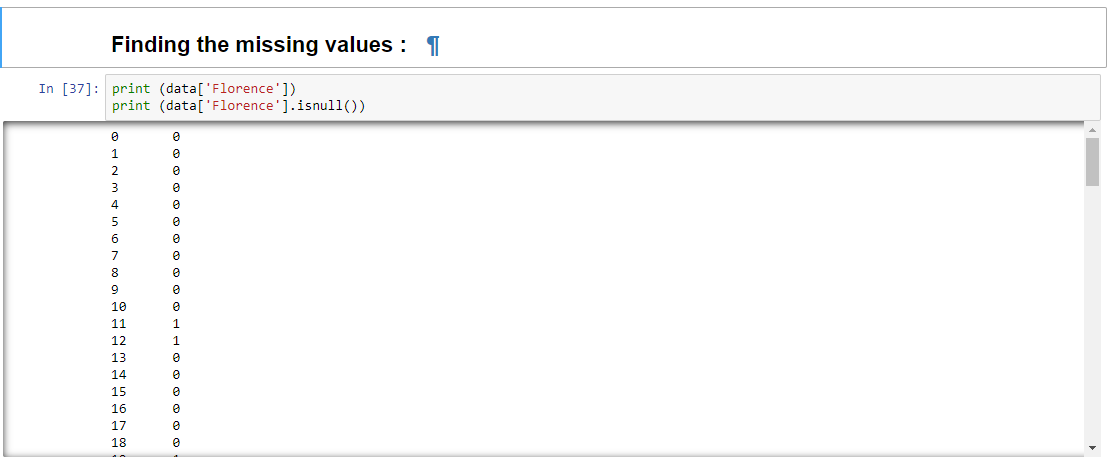
****

**DROPPING COLUMNS**

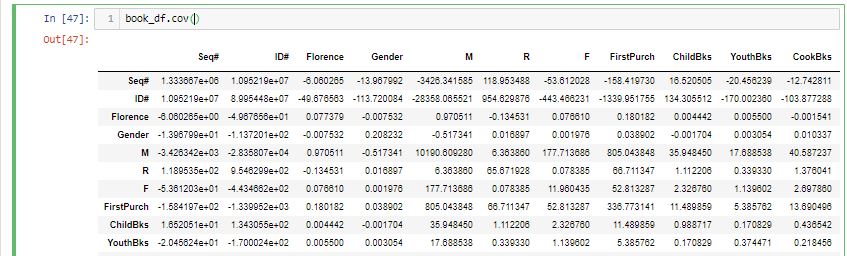
In order to drop the columns which are not further used in the data set we used drop() method.

****

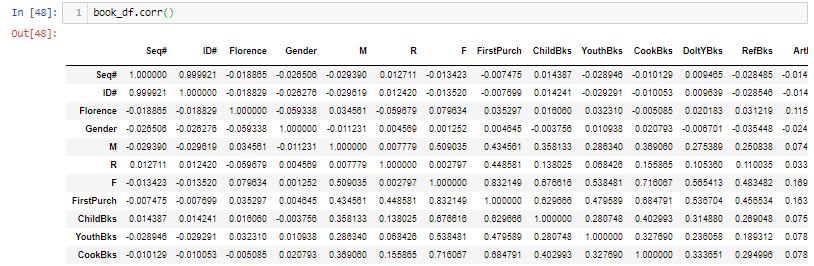
**CHECKING NULL VALUES AND MISSING VALUES**

****

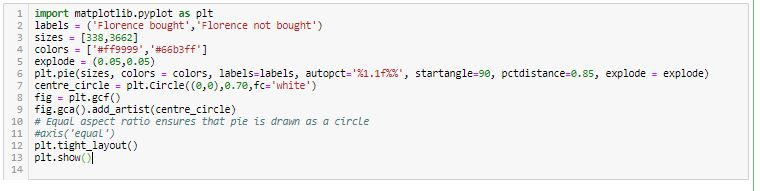
**Cov ()** function. Covariance provides the a measure of strength of correlation between two variable or more set of variables. The covariance matrix element Cij is the covariance of xi and xj.

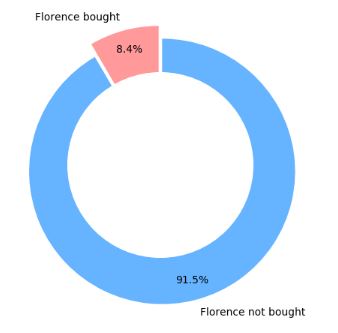


**Corr ()** is used to find the pairwise correlation of all columns in the data

****

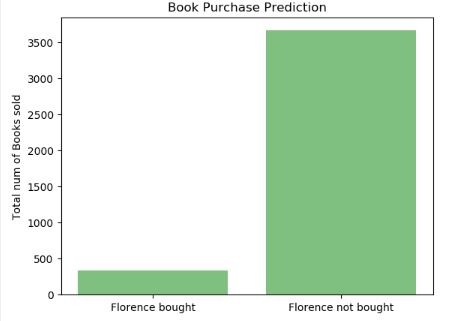
**4.1.1 PIE CHART**

****

****

**4.1.2 BAR CHART**

****

****

**4.1.3 HEAT MAP**

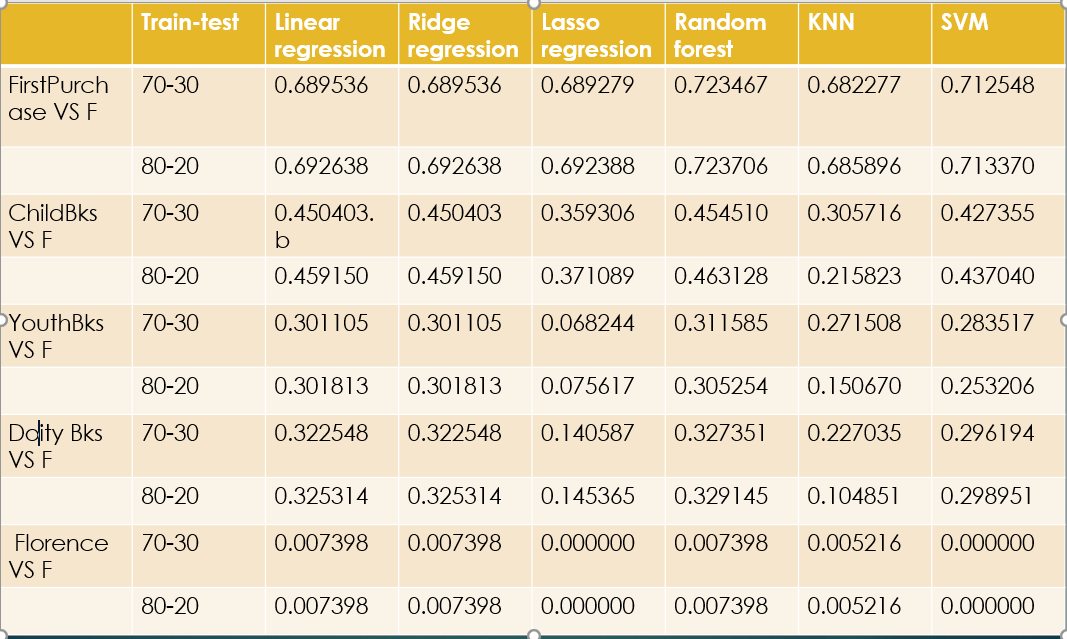
****

**5)Findings and Suggestions**:

* According to the given dataset, we found that the cookbooks are sold more at end of the period of time, so the same needs to be restocked more and ItalAtlas are sold less than everyone
* Female ratio of buying books are more than males
* Maximum number of people didn’t buy the book named florence

To solve the problem, first we generate a Florence-ID list from the original list of review information. Then we reconstruct the training and validation data by randomly selecting Florence-ID pairs which do not show up in the training set as non-purchase pairs. As for models, we considered models: logistic regression, linear regression, Lasso, Random Forest, Ridge, KNN, SVM. For popularity model, we considered the most popular pairs as LINEAR, RANDOM FOREST AND RIDGE ALGORITHMS.

This problem is useful and practical in the real life, but since the dataset is quite large, working on the whole dataset is time-consuming, maybe we should come up with some other more efficiency and effective models.

****

**6)Conclusion**:

This project is a recommender system problem, the main idea is to evaluate people’s preference and make some recommendations. In this specific problem, we are given a dataset on book prediction, which provided us purchase information, our target is to predict whether maximum number of people are willing to buy the book Florence or not.

7.BIBLIOGRAPHY AND REFERENCE:

1.International Research Journal of Engineering and Technology (IRJET)

Customer buying Prediction Using Machine-Learning Techniques: A Survey Mr. Shrey Harsh Baderiya1, Prof. Pramila M. Chawan2

<https://www.irjet.net/archives/V5/i10/IRJET-V5I10174.pdf>

2. <https://github.com/KuangyiYang/P5-Amazon-Book-Purchase-Prediction>

3. <http://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html>