Assignment One (A1):
Information Visualisation (Individual)
Pallav Shukla – 2154638 – MSc Data Science – HCVA – CSCM27

Abstract:

Visualisation of the Dry Bean dataset [1] through the process of selecting a user and task, then making a couple of prototype designs and implementing the selected design through python Altair.

Dataset Description:

There are numerous samples of dry beans, out of which seven were the most useful and important. These beans were visualised for analysing their distinct features, and a total of 16 features, 12 dimensions and four different shape forms were found [6] [2]. The dataset comprises attributes like Area, Perimeter, Solidity etc., in a total of 17 attributes with multivariate characteristics.

Defining User and Task:

Our main public includes Supermarket Giants companies that produce and process [1] the grains with the involvement of farmers. For the processing of grain [7], we need to take certain factors into consideration that boost both marketings as well as production. So, in our study, we found some co-relation that can boost our tasks in accordance with certain attributes like Solidity and Area of the grain.

Visual Design and Description:

Prototype 1 and Prototype 2 describe different simulation of Visual design done through Tableau.
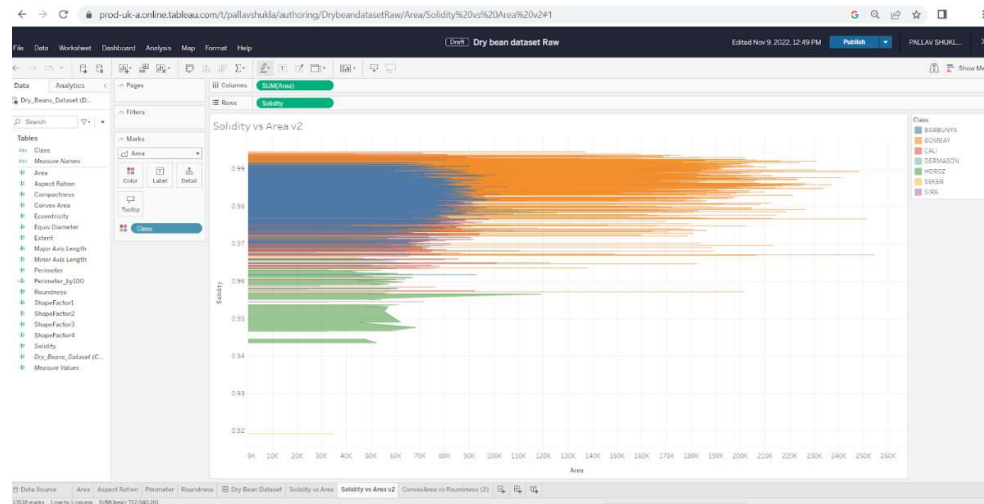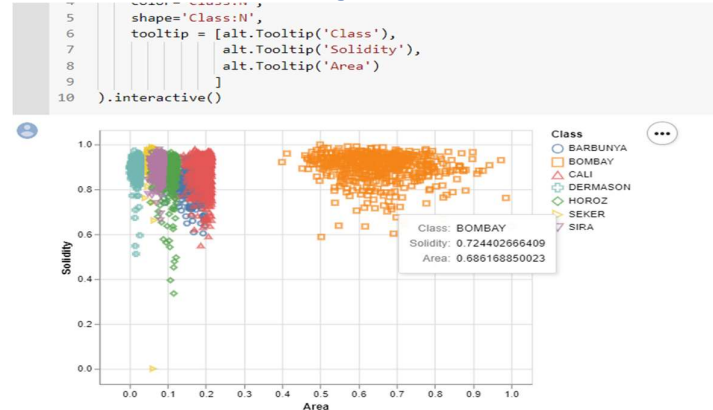

Prototype 1

Prototype 2

We created two different prototypes with different colour simulations that are parted from each other with spatial distance, shape, and colour. That helps to identify the class of grains which will yield profit to the company in accordance with production, marketing as well as storage. Out of several attributes, Solidity and Area resulted in forming a clear advantageous result. Among seven classes, two of them resulted in winning the first and second positions according to our factors considered. Class Bombay and class Dermason these classes clearly identify themselves as a clear winner in the visualisation formed, the former displaying between 130k to 230k in the majority Area wise as well as in solidity, majorly lying between .970 to .995, whereas the latter lying between 20k to 40k as well as .981 to .993 with respect to Area and Solidity respectively.

Prototype 1 has a clear benefit of representing all the elements of the specific attribute, whereas Prototype 2 has the advantage of making a clear vision of the Attributes and representing them in a user-friendly way, i.e., that is easy to understand.

Implementation and intended understanding:

Implementation of Prototype 2 design

From the above two prototypes, the prototype 2 was selected as it had a clear understanding of the concept of visualisation, as it represents the data in varying the position, 2length, spatial distances, shapes, and colours. Leading us to the concept of understanding the task according to our problem.

The user receives better knowledge as to how to progress, to earning a profit according to the graph From the above implementation result, it is clearly indicated that Bombay clearly takes up a lot of space/area but has good solidity that results in better profit in terms of holding the stock for a longer period whereas Dermason has an advantage over the feature of Area as it takes very less space so can be stored more in number but must be utilised very frequently because of its Solidity values.

In our implementation, we also included more relations that can be considered as a factor for the profit of the company, but as per our problem, that is the profit related to Marketing, production, and Storage, the solidity and area gives us a better approach.

References

1. Słowiński, G. (2021). *Dry Beans Classification Using Machine Learning*. http://ceur-ws.org
2. Dry Bean Dataset. Www.kaggle.com. Retrieved November 9, 2022, from https://www.kaggle.com/datasets/sansuthi/dry-bean-dataset
3. *UCI Machine Learning Repository: Dry Bean Dataset Data Set*. Archive.ics.uci.edu. https://archive.ics.uci.edu/ml/datasets/Dry+Bean+Dataset
4. Dry Beans Classification and Intro for Competition 3. Www.youtube.com. Retrieved November 9, 2022, from https://www.youtube.com/watch?v=jhmlsuXqrao
5. ML-Final_Project. Www.youtube.com. Retrieved November 9, 2022, from https://www.youtube.com/watch?v=b8rakETpErs
6. Watch Me Viz - #MakeoverMonday 2021 Week 14 - Multiclass Classification of Dry Beans. Www.youtube.com. Retrieved November 9, 2022, from https://www.youtube.com/watch?v=hnO2EoGneB8&t=1592s

7. ITS665 DRY BEAN DATASET. Www.youtube.com. Retrieved November 9, 2022, from https://www.youtube.com/watch?v=Zcm14hB6hMw

8. UCI Machine Learning Repository. (2018). Uci.edu. http://archive.ics.uci.edu/ml/index.php

9. Hospodarenko, H., Liubich, V., & Novikov, V. (2018). OPTIMIZATION OF PRODUCTION PROCESS OF PEELED GRAINS OF WHEAT OF DIFFERENT SOLIDITY. *EUREKA: Life Sciences*, *5*, 3–14. https://doi.org/10.21303/2504-5695.2018.00718