# DEODORANT LIKING

Submission :
Prachi Prakash

# PROBLEM STATEMENT

1. This project was done for a leading consumer packaged good company

2. The project aimed at predicting the Instant Liking of five major deodorant products namely – Deodorant B, Deodorant G, Deodorant J, Deodorant F and Deodorant H based on a set of questionnaires and responses from different consumers

3. The model formulation and predicted result consequently will help the company take informed strategic decisions

# APPROACH

The following approach was taken in order to analyze the data and derive predictions

1. Data Gathering
2. Data Cleaning
3. Data Analysis
4. Model Building
5. Model Evaluation

# DATA GATHERING

1. Data gathering or data collection is a process of gathering and measuring information which enables us to answer relevant questions and evaluate outcomes.
2. The data set for the analysis was provided in terms of comma separated values(csv) files for five deodorant products.
3. Data dictionary file was also provided along with the data set. This file helped us to gather additional information for the columns of the data set whose column names may not be self explanatory

# DATA CLEANING

Certain aspects of Data Cleaning done -

a) Question 8 – "If this fragrance had a texture, would it be" was non uniformly distributed which had to be made uniform. It was handled by creating missing attributes from q8.1 to q8.20. The columns was then converted from wide to long format for all five deodorant data sets

b) Question s13 – "Would you please tell me the brands perfume/fragrances you have worn regularly in the past 6 months?". The responses to this question was also non uniformly distributed across all data frames. Those were also treated the same way as above

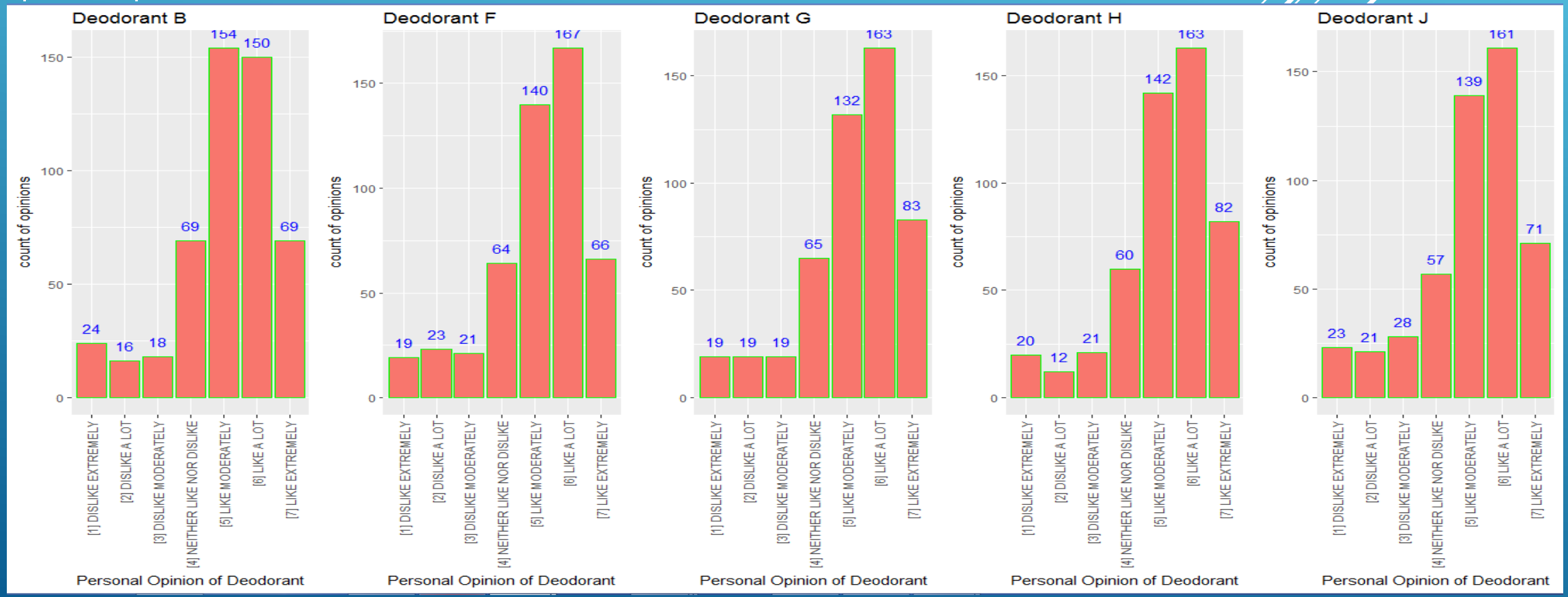
c) All the five individual deodorant data sets columns were made uniform and then merged into one master data frame for further analysis

d) Other cleaning included, converting the list data types to numeric and treating NA values

# DATA ANALYSIS

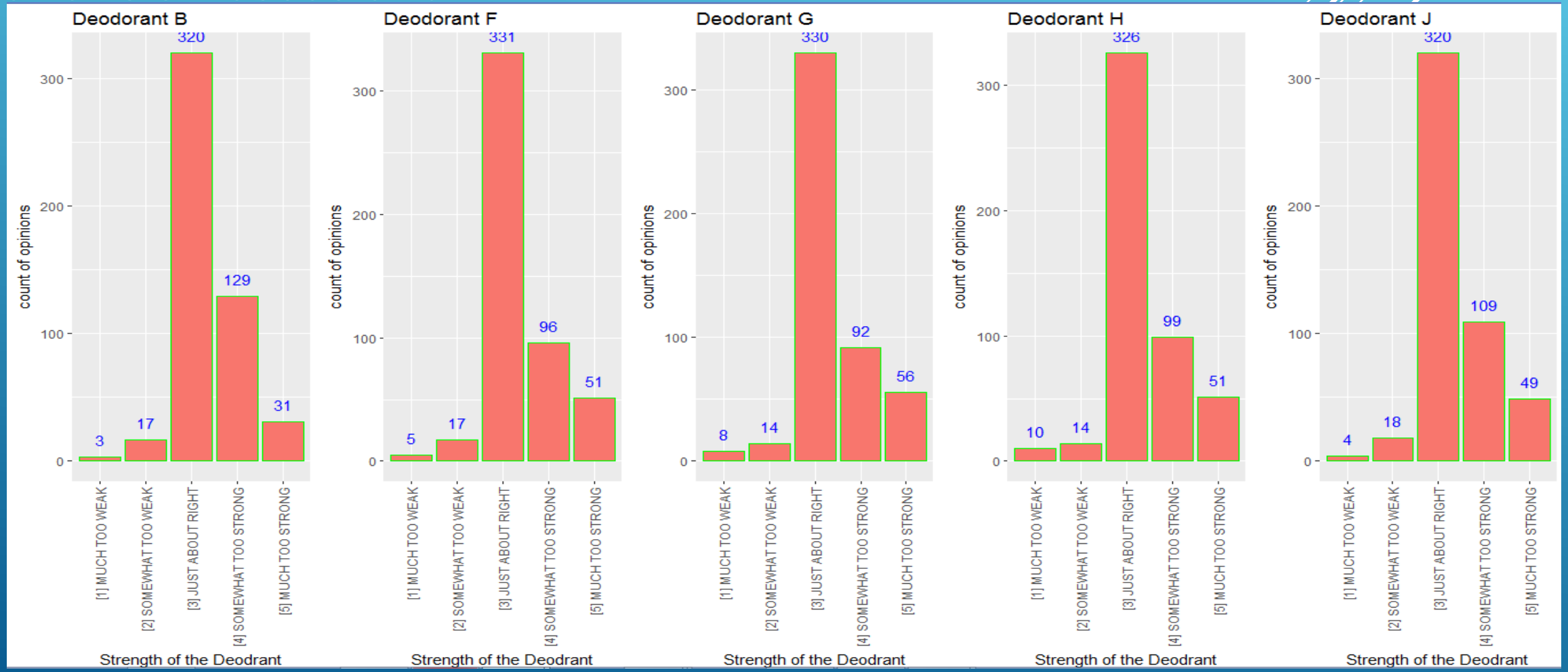
Certain Exploration in the data was done for analysis

1. Based on Personal Opinion on Deodorants, below analysis shows consumer perception on different deodorants



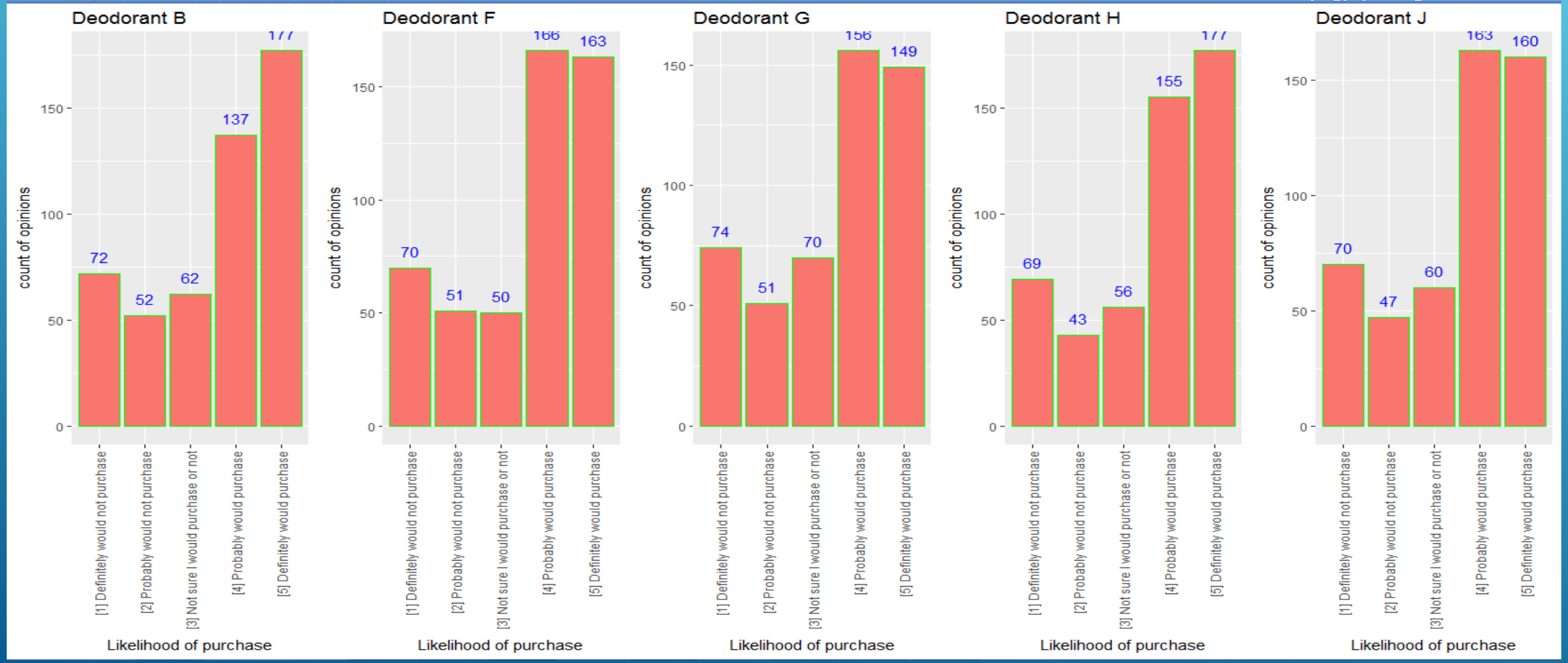
# DATA ANALYSIS

2. Based on Strength of the deodorant, below analysis shows consumer perception on different deodorants



# DATA ANALYSIS

3. Based on Likelihood of purchase, below analysis shows consumer perception on different deodorants



# MODEL BUILDING

1. In a pre model building stage, necessary columns of a master data frame (containing data for all deodorants), was converted into factors
2. Dummy variables were created where each level of categorical variable is contrasted to a specified reference level.
3. The master data was then divided into training and testing data set. Training data set was used for building a model. Testing data set was used to evaluate the model created
4. Model was build using decision tree based algorithm. It is one of the best and mostly used supervised learning method. Tree based methods empower predictive models with high accuracy, stability and ease of interpretation

# MODEL BUILDING – WHY DECISION TREE

1. The number of columns of the master data frame containing all deodorants data set were too many to use the logistic regression model. Based on the system constraints, the glm function for feature selection was taking too much time to execute
2. Feature selection is important in analytics. When we fit a decision tree on a training data set, top few nodes on which the tree is split are essentially the most important variables within the data set and feature selection is completed automatically
3. Decision tree does not make any assumptions of linearity in the data.

# THANK YOU