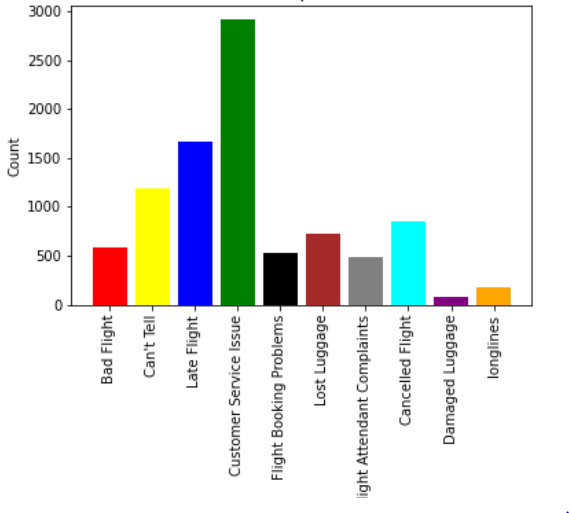
**Team mates**

1. Shivangi Verma
2. Prashant Chaudhary
3. Anirban Kar Chaudhuri
4. Ankeit Taksh
5. Marcus Yatim

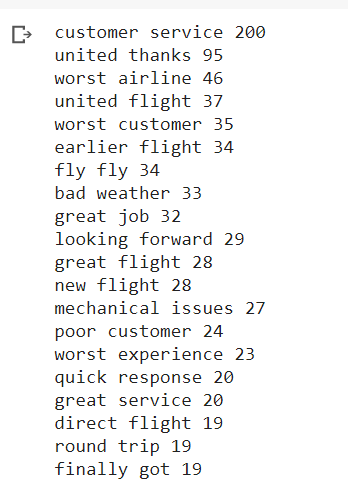
Part 1

i)Identify what aspects are of interest to airline:

Counting the data using negative reason column



After cleaning and frequency count customer service and flight issues appear to have the highest frequency



ii)Create a mapping table

aspect = {"Flight": ["flight delayed","was cancelled flightled","late flightr","late flight","cancelled flightled","flight delayed","flight cancelled","cancelled flighted","great flight"],

          "Service": ["customer service","worst customer service","poor customer service","terrible customer service","great customer service","horrible customer service","amazing customer service","excellent customer service","awful customer service","bad customer service"]}

iii)Do sentiment analysis

iv)Create a visualization

v)Identify which airline has performed best in which aspect,

Challenges faced and lesson learnt

1. Data cleaning as many tweets were having junk data or not much useful data
2. Brainstorm on how to create domain dictionary and deriving aspects
3. As many tweets had spelling mistakes and gramatiical errors, we kept them intact and also enriched the dictionary with the right phrases such that it is also useful for future tweets data for example "late flightr" and “late flight” both were kept
4. As people tend to write tweets more when they are unhappy hence we can see tweets data is moe included towards negative side
5. **The distribution of sentiment according to flight was working but runtime crash was occurring a lot of time**

Code snippet:

#Mapping aspects

# aspect categories

# From Frequency approach

aspect = {"Flight": ["flight delayed","was cancelled flightled","late flightr","late flight","cancelled flightled","flight delayed","flight cancelled","cancelled flighted","great flight"],

          "Service": ["customer service","worst customer service","poor customer service","terrible customer service","great customer service","horrible customer service","amazing customer service","excellent customer service","awful customer service","bad customer service"]}

#aspect = {"Love": ["love","Love"]}

#aspect =  {"Flight": ["flight delayed","was cancelled flightled","Late Flight"]

          # ,"Service": ["customer service"]

           #}

flight\_names=['US Airways','United','Southwest', 'Delta','American','@VirginAmerica']

# Let's also use a category 'overall' to cover everything else

# Helper function for categorising aspects:

# Here the input is output from Spacy

def aspectCat(sentence):

    category=[]

    for key, val in aspect.items():

       for t in val:

          #print("t=" + t + ", sentence.text:"+sentence.text )

          if t.lower() in sentence.text.lower(): category.append(key)

    if len(category) == 0 : category.append("overall")

    return category

def flightName(sentence):

    f='none'

    for flight in flight\_names:

       if flight.lower() in sentence.text.lower(): f=flight.lower()

    return f