

Twitter US Airline Sentiment

INTRODUCTION

Analyze how travelers in February 2015 expressed their feelings on Twitter

In this phase the design to innovation and data flow of customer segmentation is going to be done.

DATASET

The data is obtained from [//www.kaggle.com/datasets/crowdflower/twitter-airline-sentiment](http://www.kaggle.com/datasets/crowdflower/twitter-airline-sentiment)

COLUMNS USED

From Tweets.csv data the following columns are used

- tweet_id
- airline_sentiment
- negativereason
- airline
- name

LIBRARIES USED

The essential libraries used in this project are :

- RSQLite
- Dplyr
- Ggvis
- Wordcloud

I've done some minimal preprocessing on this data and re-released it on [Kaggle](#) as a CSV file and SQLite database.

```
library(RSQLite)
```

```
db <- dbConnect(dbDriver("SQLite"), "../input/database.sql")
```

First, let's see what tables we have to work with.

```
library(dplyr)
```

```
tables <- dbGetQuery(db, "SELECT Name FROM sqlite_master WHERE type='table'")
```

```
colnames(tables) <- c("Name")
```

```
tables <- tables %>%
```

```
  rowwise() %>%
```

```
  mutate(RowCount=dbGetQuery(db, paste0("SELECT COUNT(*) RowCount FROM ", Name)))$RowCount[1])
```

```
print.table(tables)
```

As we see above, there's a single table: Tweets. Now let's see what this table contains.

```
print.table(dbGetQuery(db, "
```

```
SELECT *
```

```
FROM Tweets
```

```
LIMIT 6"))
```

```
print.table(dbGetQuery(db, "
```

```
SELECT airline,
```

```
  negativereason,
```

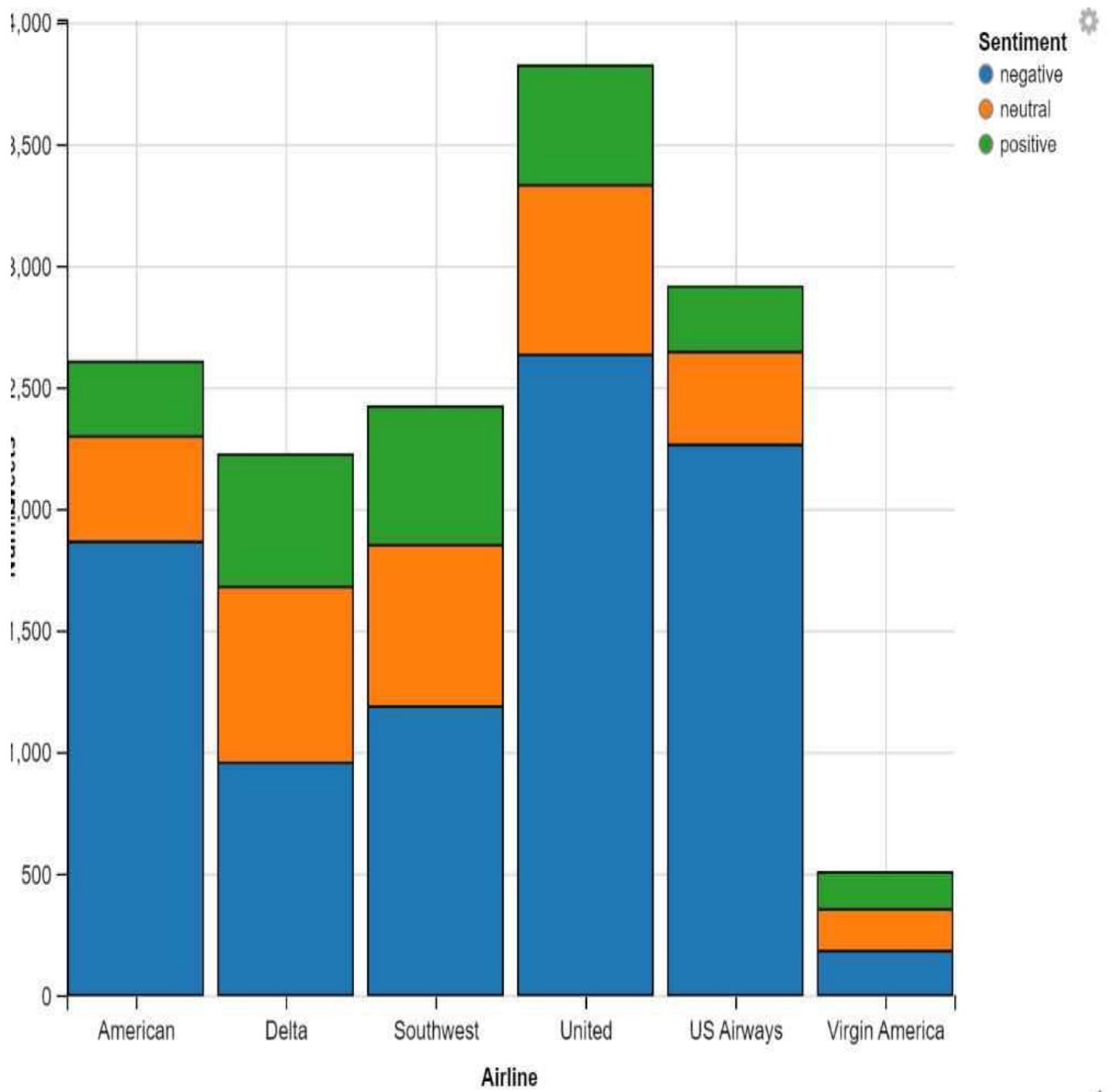
```
  COUNT(negativereason)
```

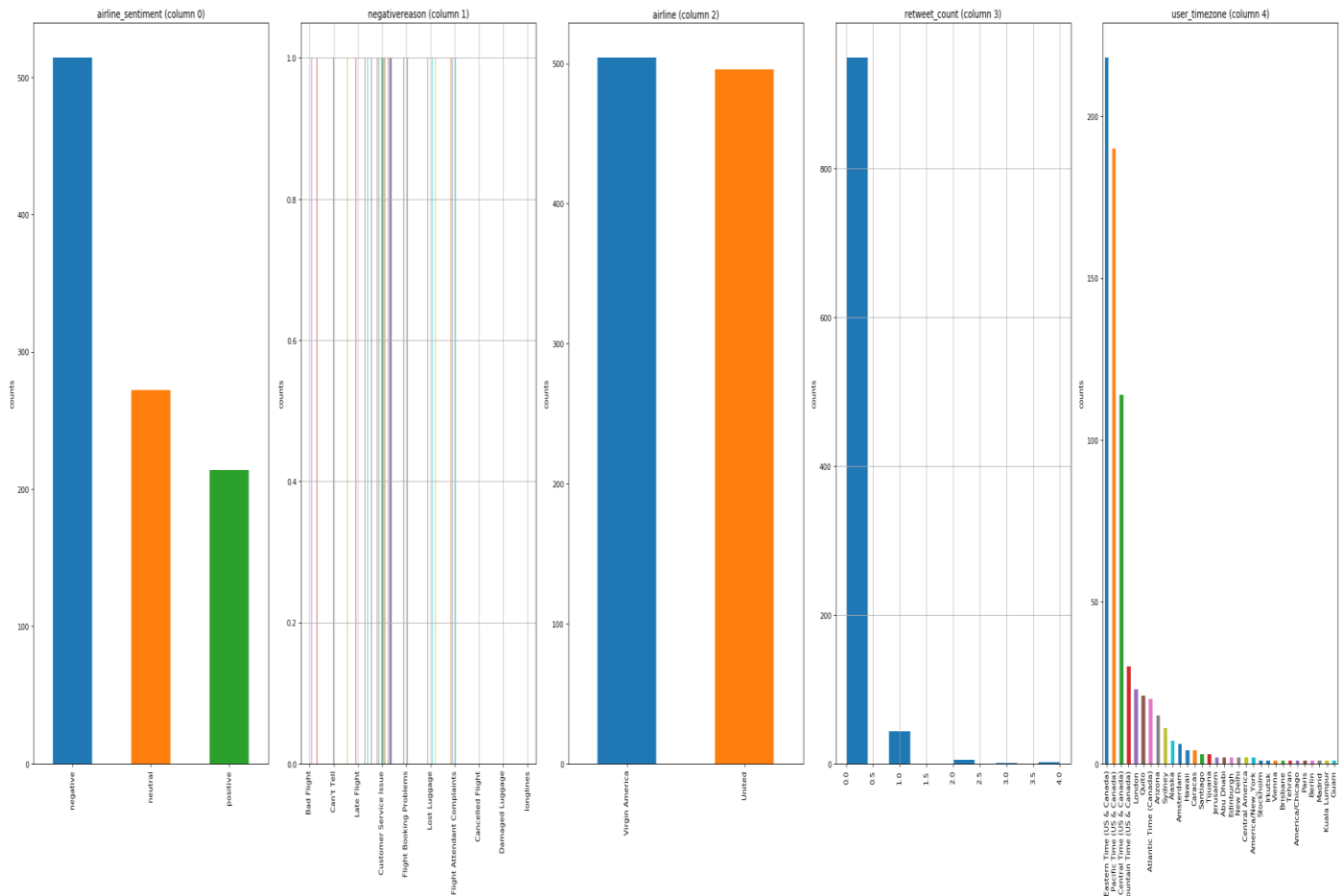
```
FROM Tweets
```

```
GROUP BY airline,
```

```
  negativereason
```

```
ORDER BY COUNT(negativereason) DESC"))
```





Exploratory Analysis

To begin this exploratory analysis, first use `matplotlib` to import libraries and define functions for plotting the data. Depending on the data, not all plots will be made. (Hey, I'm just a kerneling bot, not a Kaggle Competitions Grandmaster!)

```
from mpl_toolkits.mplot3d import Axes3D
from sklearn.preprocessing import StandardScaler
import matplotlib.pyplot as plt # plotting
import numpy as np # linear algebra
import os # accessing directory structure
import pandas as pd # data processing, CSV file I/O (e.g. p
d.read_csv)
```

There is 1 csv file in the current version of the dataset

Conclusion

This concludes your starter analysis! To go forward from here, click the blue "Edit Notebook" button at the top of the kernel. This will create a copy of the code and environment for you to edit. Delete, modify, and add code as you please.