ola-vs-uber

April 30, 2024

[1]:
```
# This Python 3 environment comes with many helpful analytics libraries␣
 ↪installed
# It is defined by the kaggle/python Docker image: https://github.com/kaggle/
 ↪docker-python
# For example, here's several helpful packages to load

import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)

# Input data files are available in the read-only "../input/" directory
# For example, running this (by clicking run or pressing Shift+Enter) will list␣
 ↪all files under the input directory

import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))

# You can write up to 20GB to the current directory (/kaggle/working/) that␣
 ↪gets preserved as output when you create a version using "Save & Run All"
# You can also write temporary files to /kaggle/temp/, but they won't be saved␣
 ↪outside of the current session
```

```
/kaggle/input/ola-vs-uber-play-store-reviews/Ola Customer Reviews.csv
/kaggle/input/ola-vs-uber-play-store-reviews/Uber Customer Reviews.csv
```

1 Loading Data

[2]:
```
Ola = pd.read_csv('/kaggle/input/ola-vs-uber-play-store-reviews/Ola Customer␣
 ↪Reviews.csv')
Uber = pd.read_csv('/kaggle/input/ola-vs-uber-play-store-reviews/Uber Customer␣
 ↪Reviews.csv')
```

```
/tmp/ipykernel_20/63859973.py:1: DtypeWarning: Columns (3) have mixed types.
Specify dtype option on import or set low_memory=False.
  Ola = pd.read_csv('/kaggle/input/ola-vs-uber-play-store-reviews/Ola Customer
Reviews.csv')
```

```
/tmp/ipykernel_20/63859973.py:2: DtypeWarning: Columns (3) have mixed types.
Specify dtype option on import or set low_memory=False.
  Uber = pd.read_csv('/kaggle/input/ola-vs-uber-play-store-reviews/Uber Customer
Reviews.csv')
```

2 EDA - Analysis

[3]:
```
Ola
```

[3]:
```
             source                             review_id         user_name  \
0       Google Play  fbc7ffc9-5a89-446e-87fd-d69bf4a7f984     Puipuii Ralte
1       Google Play  5a0051fb-220a-45b2-ba94-a15a2949218f      Deepak Kumar
2       Google Play  71ebf933-b734-474d-bb65-a18c90906ed2  Ahamed Azarudeen
3       Google Play  e1cc0010-60b3-4126-99c2-e8549088566a        Rahil Syed
4       Google Play  77cf1be1-b428-4493-ae25-e0f288f79b8f           vin 007
…             …                                    …               …
357693    App Store  575258ed-aec1-47ea-b792-88deb17e4ad7          Jayken17
357694    App Store  ca91ebc0-92d9-48a6-9c46-7dcad1a0546e       cbarath1986
357695    App Store  f7227b64-90aa-4c82-996e-c86d99761831          MaS Mitt
357696    App Store  5f14c66e-94cc-4594-a83e-055ab8721ca8         vasantha2
357697    App Store  b4f2a564-e5aa-48d5-bbc0-8b8e2751628e          June Day

                          review_title  \
0                                  NaN
1                                  NaN
2                                  NaN
3                                  NaN
4                                  NaN
…                                  …
357693  Can't sign up with App200 code
357694     Worst app n online services
357695   Signup itself is so difficult
357696                Worst iphone App
357697                        Hopeless

                                    review_description  rating  thumbs_up  \
0       The map in Ola is so messed up, i have to pay …       1        0.0
1                                 Deepak Kumar…    ]       5        0.0
2       Such aa irresponsible app more then I waiting …       1        0.0
3                                                 Worst       1        0.0
4       Too much expensive .. try UBer… They are pro…       1        0.0
…                                                   …     …        …
357693  The app is useless for new users since you can…       1        NaN
357694  Do not recharge money with ola money.. I have …       1        NaN
357695  The app hangs on signup. Later I get authentic…       2        NaN
357696  one of the worst app for iphone.i was tried mu…       1        NaN
357697  This company is hopeless. I waited on hold for…       1        NaN
```

```
                review_date developer_response developer_response_date  \
0       2023-08-10 16:40:50                NaN                      NaN
1       2023-08-10 16:36:14                NaN                      NaN
2       2023-08-10 16:29:31                NaN                      NaN
3       2023-08-10 15:52:06                NaN                      NaN
4       2023-08-10 15:51:10                NaN                      NaN
…                       …                  …                        …
357693  2015-03-08 04:33:32                NaN                      NaN
357694  2015-03-02 23:13:22                NaN                      NaN
357695  2015-03-02 11:52:50                NaN                      NaN
357696  2015-02-17 15:17:12                NaN                      NaN
357697  2014-10-16 08:53:01                NaN                      NaN

       appVersion laguage_code country_code
0           6.3.2           en           in
1             NaN           en           in
2           6.3.1           en           in
3           5.0.4           en           in
4             NaN           en           in
…             …             …            …
357693        NaN           en           in
357694        NaN           en           in
357695        NaN           en           in
357696        NaN           en           in
357697        NaN           en           in

[357698 rows x 13 columns]
```

[4]:
```
Uber
```

[4]:
```
              source                             review_id          user_name  \
0        Google Play  18d6584c-d0e9-4833-a744-f607058aee97          Milky Way
1        Google Play  50a08f18-cece-4ddf-b617-028844c8aa28     Bradlee Severa
2        Google Play  b0d8e75a-80a7-4dcd-abaf-72b046dbeeb7      Amit Aggarwal
3        Google Play  502702a9-25ed-4373-a96c-7fa1f06caacd       Bryant Inman
4        Google Play  f47a3fb6-23db-49bd-9e63-f33c8d724d07    Addie Whittaker
…                …                                    …                 …
1069611    App Store  015547c9-1d97-4b92-8206-ef47a540b70b  Ad hater 20140323
1069612    App Store  e1125a24-a804-419e-8aa2-039e3f380d25        valeramos02
1069613    App Store  132aac5d-10df-4207-a71d-01d81a4efde0        Janeé Brown
1069614    App Store  99864769-f3f9-49fc-841e-3230a72fe18e        zachwiesler
1069615    App Store  93f3188d-db2e-4532-bde3-6ec432558b5b       formerbaker1

                                          review_title  \
0                                                  NaN
1                                                  NaN
```

```
2                                                       NaN
3                                                       NaN
4                                                       NaN
…                                                       …
1069611                                       Map problems
1069612                                  Quality decrease
1069613  Uber pool walking blocks to get to the ride is…
1069614                           TERRIBLE CUSTOMER SERVICE
1069615                               Poor Customer Service

                                      review_description  rating  thumbs_up  \
0        Suddenly, the driver can't have my location an…       1        0.0
1        Very cordial.. And helped with a quick turnaro…       5        0.0
2                                    Very good experience       5        0.0
3                                             All I use         5        0.0
4        I have enjoyed traveling by Uber my drivers ha…       5        0.0
…                                                       …     …        …
1069611  I tried to find away to report problems direct…       3        NaN
1069612  I used to love Uber, specially the Uber pool s…       2        NaN
1069613  If I wanted to take a bus to be dropped off on…       1        NaN
1069614  Hello\n\nSTORY TIME\n\nI wanted to delete JUST…       1        NaN
1069615  The past couple times I've ridden with Uber, I…       1        NaN

                review_date  \
0        2023-08-10 17:48:51
1        2023-08-10 17:38:35
2        2023-08-10 17:38:17
3        2023-08-10 17:37:45
4        2023-08-10 17:36:56
…                        …
1069611  2019-02-20 19:41:25
1069612  2019-01-24 12:13:39
1069613  2019-01-14 16:53:31
1069614  2019-01-06 17:19:59
1069615  2019-01-06 12:34:01

                                      developer_response  \
0                                                       NaN
1                                                       NaN
2                                                       NaN
3                                                       NaN
4                                                       NaN
…                                                       …
1069611  {'id': 7416266, 'body': "Hi, this doesn't soun…
1069612  {'id': 7006213, 'body': "Hi, uber works better…
1069613  {'id': 6859976, 'body': 'Hi Janee, this certai…
1069614                                                 NaN
```

```
1069615  {'id': 6709611, 'body': 'Hi, this certainly so…

        developer_response_date   appVersion laguage_code country_code
0                           NaN          NaN           en           in
1                           NaN  4.485.10000           en           in
2                           NaN  4.486.10002           en           in
3                           NaN  4.467.10008           en           in
4                           NaN  4.486.10002           en           in
…                           …            …            …            …
1069611                     NaN          NaN           en           in
1069612                     NaN          NaN           en           in
1069613                     NaN          NaN           en           in
1069614                     NaN          NaN           en           in
1069615                     NaN          NaN           en           in

[1069616 rows x 13 columns]
```

[5]:
```
Ola.shape
```

[5]:
```
(357698, 13)
```

[6]:
```
Uber.shape
```

[6]:
```
(1069616, 13)
```

[7]:
```
Ola.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 357698 entries, 0 to 357697
Data columns (total 13 columns):
 #   Column                   Non-Null Count   Dtype
---  ------                   --------------   -----
 0   source                   357698 non-null  object
 1   review_id                357698 non-null  object
 2   user_name                357698 non-null  object
 3   review_title             891 non-null     object
 4   review_description       357682 non-null  object
 5   rating                   357698 non-null  int64
 6   thumbs_up                356807 non-null  float64
 7   review_date              357698 non-null  object
 8   developer_response       124769 non-null  object
 9   developer_response_date  124590 non-null  object
 10  appVersion               275326 non-null  object
 11  laguage_code             357698 non-null  object
 12  country_code             357698 non-null  object
dtypes: float64(1), int64(1), object(11)
memory usage: 35.5+ MB
```

[8]:
```
Uber.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1069616 entries, 0 to 1069615
Data columns (total 13 columns):
 #   Column                  Non-Null Count    Dtype
---  ------                  --------------    -----
 0   source                  1069616 non-null  object
 1   review_id               1069616 non-null  object
 2   user_name               1069615 non-null  object
 3   review_title            2180 non-null     object
 4   review_description      1069501 non-null  object
 5   rating                  1069616 non-null  int64
 6   thumbs_up               1067436 non-null  float64
 7   review_date             1069616 non-null  object
 8   developer_response      198264 non-null   object
 9   developer_response_date 197278 non-null   object
 10  appVersion              828068 non-null   object
 11  laguage_code            1069616 non-null  object
 12  country_code            1069616 non-null  object
dtypes: float64(1), int64(1), object(11)
memory usage: 106.1+ MB
```

[9]:
```
Ola.describe()
```

[9]:

	rating	thumbs_up
count	357698.00000	356807.000000
mean	2.77325	0.883461
std	1.85194	16.417954
min	1.00000	0.000000
25%	1.00000	0.000000
50%	2.00000	0.000000
75%	5.00000	0.000000
max	5.00000	2788.000000

[10]:
```
Uber.describe()
```

[10]:

	rating	thumbs_up
count	1.069616e+06	1.067436e+06
mean	3.650441e+00	8.954054e-01
std	1.743725e+00	2.042451e+01
min	1.000000e+00	0.000000e+00
25%	1.000000e+00	0.000000e+00
50%	5.000000e+00	0.000000e+00
75%	5.000000e+00	0.000000e+00
max	5.000000e+00	5.572000e+03

3 Average Rating

[11]:
```
average_rating_ola = Ola['rating'].mean()
print("Average Rating of Ola:", average_rating_ola)
```

```
Average Rating of Ola: 2.7732500601065704
```

[12]:
```
average_rating_uber = Uber['rating'].mean()
print("Average Rating of Uber:", average_rating_uber)
```

```
Average Rating of Uber: 3.6504409058952
```

4 Preprocessing for Sentiment Analysis

[13]:
```
import matplotlib.pyplot as plt
import seaborn as sns
import nltk
import re
from nltk.corpus import stopwords
from sklearn.feature_extraction.text import TfidfVectorizer
from wordcloud import WordCloud,STOPWORDS
from nltk.stem.snowball import SnowballStemmer
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfTransformer
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import accuracy_score
from sklearn.multiclass import OneVsRestClassifier
from sklearn.svm import LinearSVC
from sklearn.linear_model import LogisticRegression
from sklearn.pipeline import Pipeline
from sklearn.preprocessing import MultiLabelBinarizer
from sklearn.metrics import f1_score
from sklearn.metrics import hamming_loss
from skmultilearn.problem_transform import BinaryRelevance
from sklearn.naive_bayes import MultinomialNB
from skmultilearn.problem_transform import ClassifierChain
from skmultilearn.problem_transform import LabelPowerset
```

```
/opt/conda/lib/python3.10/site-packages/scipy/__init__.py:146: UserWarning: A
NumPy version >=1.16.5 and <1.23.0 is required for this version of SciPy
(detected version 1.23.5
  warnings.warn(f"A NumPy version >={np_minversion} and <{np_maxversion}"
```

[14]:
```
#Remove Stopwords
stop_words = set(stopwords.words('english'))
```

```
# function to remove stopwords
def remove_stopwords(text):
    no_stopword_text = [w for w in text.split() if not w in stop_words]
    return ' '.join(no_stopword_text)

#Clean Text
def clean_text(text):
    text = text.lower()
    text = re.sub("[^a-zA-Z]"," ",text)
    text = ' '.join(text.split())
    return text

#stemming
stemmer = SnowballStemmer("english")
def stemming(sentence):
    stemSentence = ""
    for word in sentence.split():
        stem = stemmer.stem(word)
        stemSentence += stem
        stemSentence += " "
    stemSentence = stemSentence.strip()
    return stemSentence
Ola['review_description'] = Ola['review_description'].astype(str)
Ola['review_description']= Ola['review_description'].apply(lambda x:␣
 ↪remove_stopwords(x))
Ola['review_description'] = Ola['review_description'].apply(lambda x:
 ↪clean_text(x))
Ola['review_description'] = Ola['review_description'].apply(stemming)


Uber['review_description'] = Uber['review_description'].astype(str)
Uber['review_description']= Uber['review_description'].apply(lambda x:␣
 ↪remove_stopwords(x))
Uber['review_description'] = Uber['review_description'].apply(lambda x:
 ↪clean_text(x))
Uber['review_description'] = Uber['review_description'].apply(stemming)
```

[15]:
```
Ola['review_description']
```

[15]:
```
0         the map ola mess up pay rs extra map incorrect…
1                                              deepak kumar
2         such aa irrespons app i wait hour wast app ple…
3                                                    worst
4         too much expens tri uber they provid cheap rid…
                               …
357693    the app useless new user sinc can t sign refer…
```

```
357694    do recharg money ola money i recharg rs n dint…
357695    the app hang signup later i get authent code s…
357696    one worst app iphon i tri multipl time never w…
357697    this compani hopeless i wait hold forev coupl …
Name: review_description, Length: 357698, dtype: object
```

[16]:
```
Uber['review_description']
```

[16]:
```
0          sudden driver can t locat call ask i i go it s…
1          veri cordial and help quick turnaround ride we…
2                                          veri good experi
3                                                 all i use
4          i enjoy travel uber driver polit good conversa…
                                …
1069611    i tri find away report problem direct uber app…
1069612    i use love uber special uber pool servic affor…
1069613    if i want take bus drop corner walk coupl bloc…
1069614    hello stori time i want delet just drive partn…
1069615    the past coupl time i ve ridden uber i ve quit…
Name: review_description, Length: 1069616, dtype: object
```

5 Model Training for Sentiment Analysis

[17]:
```
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import CountVectorizer

# Assuming you have a DataFrame 'df' with columns 'review_description' and␣
 ↪'rating'
X = Ola['review_description']
y = Ola['rating']

# Convert ratings to binary sentiment labels (1 for positive, 0 for negative)
y = (y > 3).astype(int)

# Split data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,␣
 ↪random_state=42)

# Create a Bag-of-Words representation of the text data
vectorizer = CountVectorizer()
X_train_vectorized = vectorizer.fit_transform(X_train)
X_test_vectorized = vectorizer.transform(X_test)
```

[18]:
```
from sklearn.naive_bayes import MultinomialNB

# Create and train the model
```

```
model = MultinomialNB()
model.fit(X_train_vectorized, y_train)
```

[18]: MultinomialNB()

6 Classification Report

[19]:
```
from sklearn.metrics import accuracy_score, classification_report

# Make predictions on the testing data
y_pred = model.predict(X_test_vectorized)

# Evaluate the model
accuracy = accuracy_score(y_test, y_pred)
report = classification_report(y_test, y_pred)

print("Accuracy:", accuracy)
print("Classification Report:\n", report)
```

```
Accuracy: 0.8975258596589321
Classification Report:
              precision    recall  f1-score   support

           0       0.91      0.92      0.91     41008
           1       0.89      0.87      0.88     30532

    accuracy                           0.90     71540
   macro avg       0.90      0.89      0.90     71540
weighted avg       0.90      0.90      0.90     71540
```

7 Example

[20]:
```
new_text = ["This drive was amazing! Bad driver tho"]
new_text_vectorized = vectorizer.transform(new_text)
```

[21]:
```
predicted_sentiment = model.predict(new_text_vectorized)

if predicted_sentiment[0] == 1:
    sentiment_label = "Positive"
else:
    sentiment_label = "Negative"

print("Predicted Sentiment:", sentiment_label)
```

```
Predicted Sentiment: Negative
```

8 Pandas Profiler EDA

[22]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import ydata_profiling as pp
import seaborn as sns
import warnings
import os
```

```
/opt/conda/lib/python3.10/site-packages/numba/core/decorators.py:262:
NumbaDeprecationWarning: numba.generated_jit is deprecated. Please see the
documentation at:
https://numba.readthedocs.io/en/stable/reference/deprecation.html#deprecation-
of-generated-jit for more information and advice on a suitable replacement.
  warnings.warn(msg, NumbaDeprecationWarning)
/opt/conda/lib/python3.10/site-
packages/visions/backends/shared/nan_handling.py:51: NumbaDeprecationWarning:
The 'nopython' keyword argument was not supplied to the 'numba.jit'
decorator. The implicit default value for this argument is currently False, but
it will be changed to True in Numba 0.59.0. See
https://numba.readthedocs.io/en/stable/reference/deprecation.html#deprecation-
of-object-mode-fall-back-behaviour-when-using-jit for details.
  def hasna(x: np.ndarray) -> bool:
```

[23]:
```
pp.ProfileReport(Ola)
```

```
Summarize dataset:   0%|          | 0/5 [00:00<?, ?it/s]
Generate report structure:   0%|          | 0/1 [00:00<?, ?it/s]
Render HTML:   0%|          | 0/1 [00:00<?, ?it/s]
<IPython.core.display.HTML object>
```

[23]:

[ ]:
```

```