Project 5: Sentiment analysis for marketing

Project Title: *Sentiment Analysis*

PHASE 1 : Problem definition and design thinking

Problem statement:

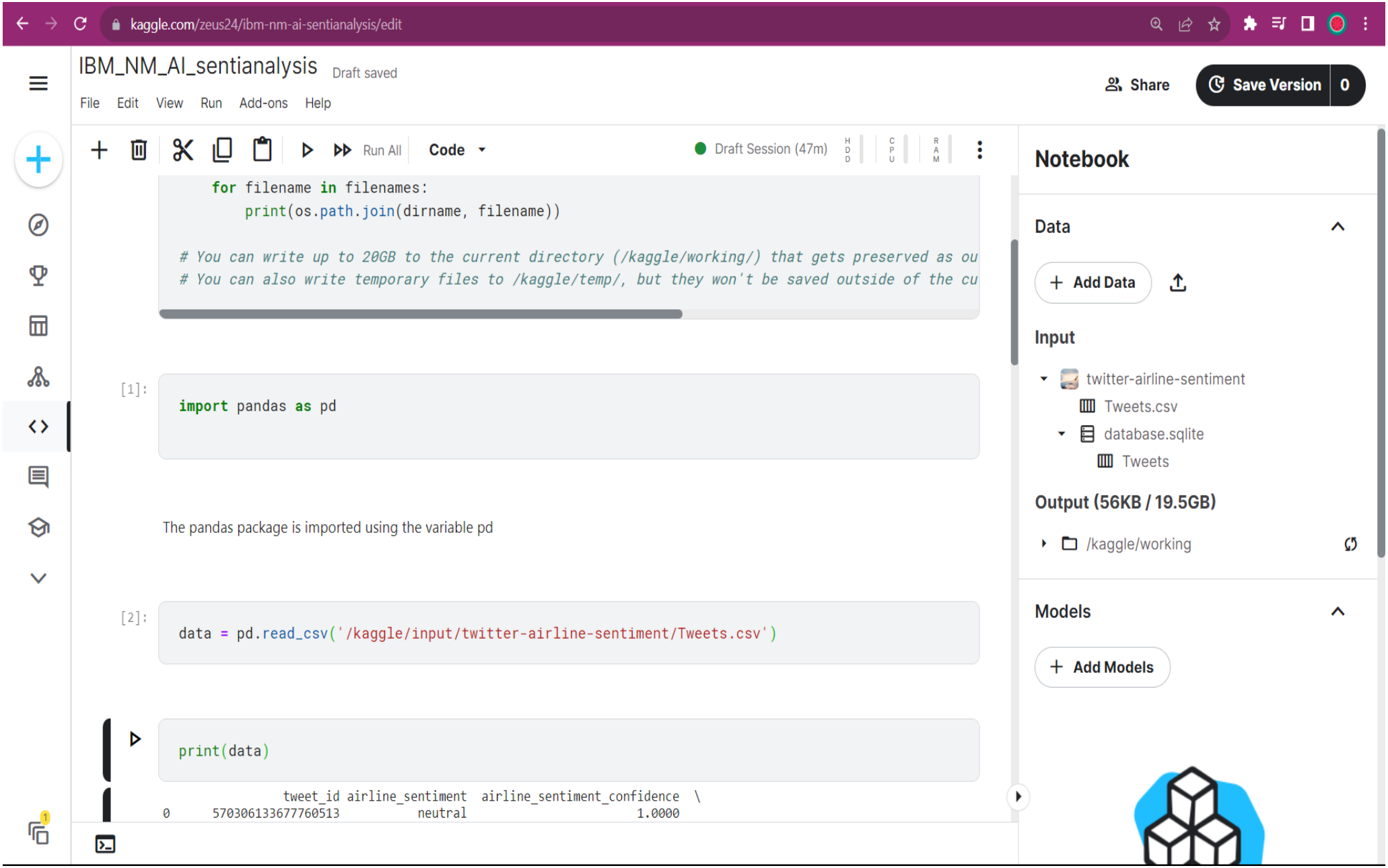
This type of project can show you what it's like to work as an NLP specialist. For this project, you want to find out how customers evaluate competitor products, i.e., what they like and dislike. It's a great business case. Learning what customers like about competing products can be a great way to improve your own product, so this is something that many companies are actively trying to do. Employ different NLP methods to get a deeper understanding of customer feedback and opinion.

Problem definition:

The problem is to perform sentiment analysis on customer feedback to gain insights into competitor products. By understanding customer sentiments, companies can identify strengths and weaknesses in competing products, thereby improving their own offerings. This project requires utilizing various NLP methods to extract valuable insights from customer feedback.

Submission by : VINOTH P G ,

Department of Electronics & Communication, Anna university RC Coimbatore.


kaggle.com/zeus24/ibm-nm-ai-sentianalysis/edit
IBM_NM_AI_sentianalysis
Draft saved
File Edit View Run Add-ons Help
Share
Save Version 0
Run All
Code
Draft Session (47m)
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 ou
# You can also write temporary files to /kaggle/temp/, but they won't be saved outside of the cu
[1]:
import pandas as pd
The pandas package is imported using the variable pd
[2]:
data = pd.read_csv('/kaggle/input/twitter-airline-sentiment/Tweets.csv')
print(data)
tweet_id airline_sentiment airline_sentiment_confidence \
0 570306133677760513 neutral 1.0000
Notebook
Data
Add Data
Input
twitter-airline-sentiment
Tweets.csv
database.sqlite
Tweets
Output (56KB / 19.5GB)
/kaggle/working
Models
Add Models

Problem statement explanation:

- The project's objective is to perform sentiment analysis on customer feedback regarding competitor products.
- By utilizing various NLP methods, the aim is to extract valuable insights into customer sentiments, allowing companies to identify strengths and weaknesses in competing products.
- These insights enable data-driven decision-making, influencing product development, marketing strategies, and ultimately, improving overall customer satisfaction and competitive advantage.
- Continuous monitoring of evolving sentiments is essential for staying responsive to changing customer perceptions.

WHAT IS SENTIMENTAL ANALYSIS?

- Sentiment analysis is a Natural Language Processing (NLP) task that involves determining the sentiment or emotional tone expressed in a piece of text(Involves understanding emotions through symbolic expressions and text data)

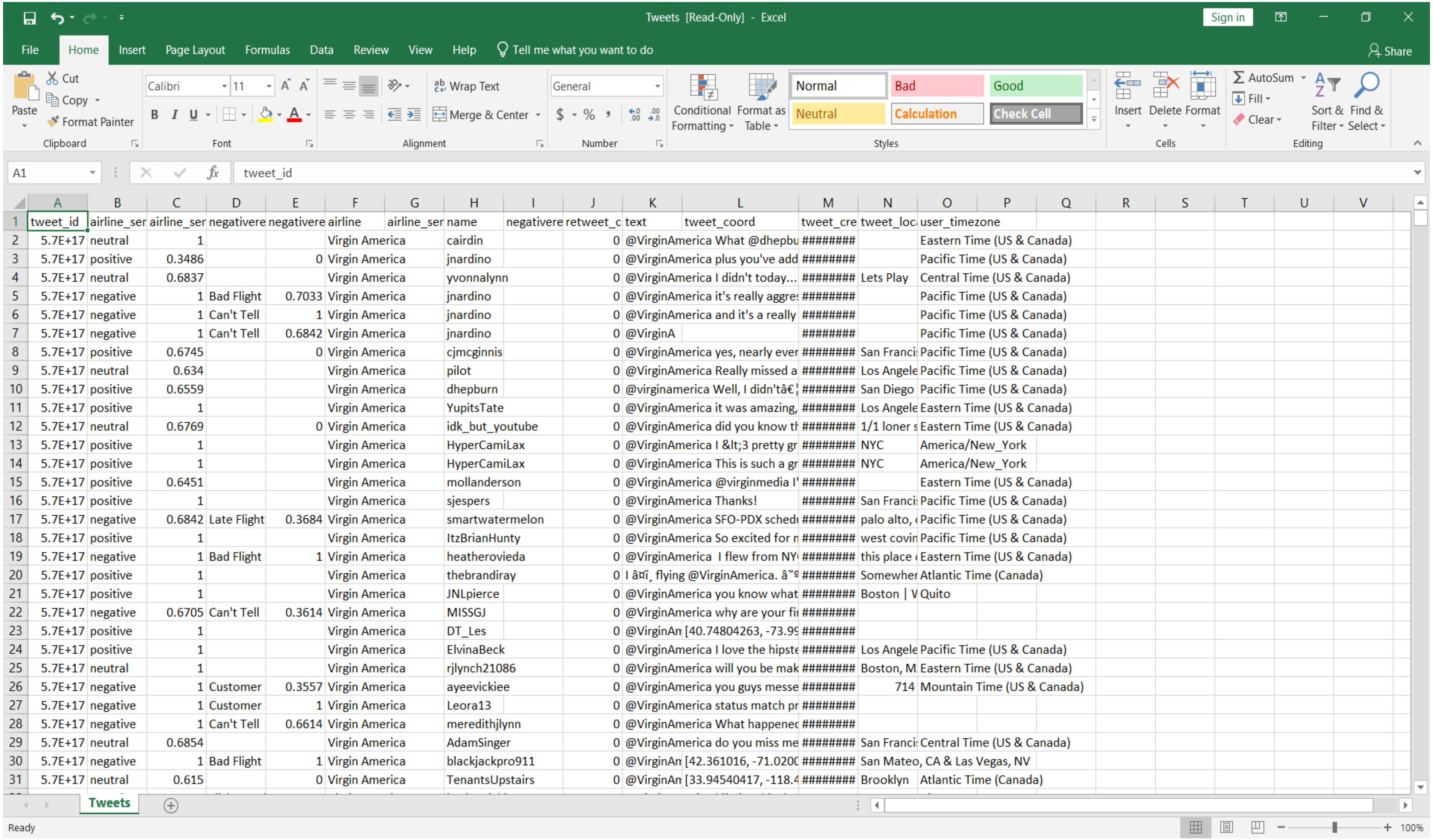
PROJECT OBJECTIVE: To perform sentimental analysis on the given dataset – **twitter US airlines sentiment** to guide the company's decision

COLLECTION OF DATASET AND PRE-PROCESSING:

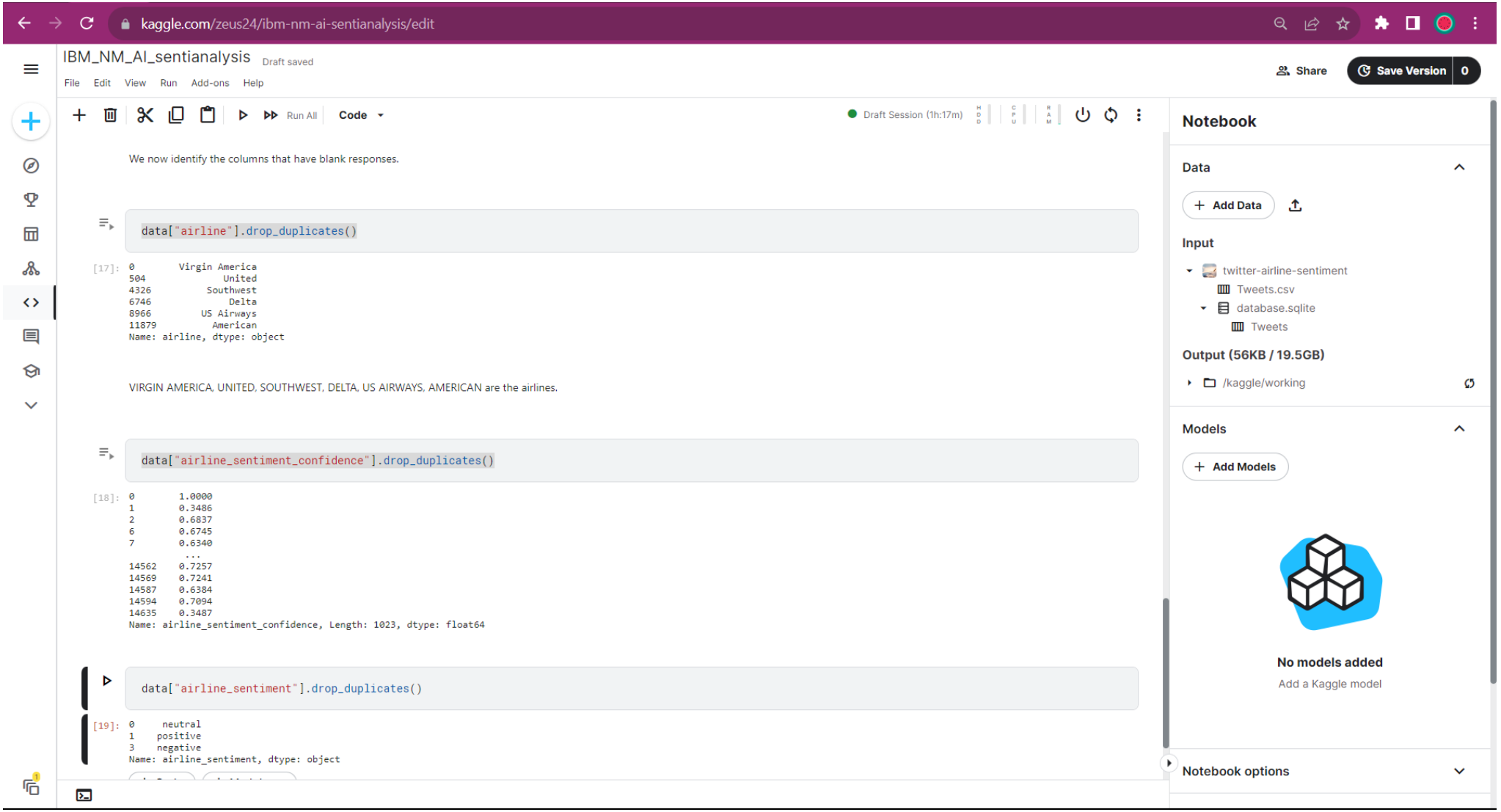
- The data is collected from the allotted dataset -- Twitter US Airline Sentiment ***{ SOURCE : KAGGLE }***
- We analyse the dataset for unfilled/empty responses.
- Either the given data can be structured into a proper form or we can deploy ML and deep learning to process the unstructured data.
- The dataset is imputated for the empty spaces using statistical techniques like –

  Mean - Primarily for numerical data

  Median and Mode – for Textual and other forms of data



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	tweet_id	airline_ser	airline_ser	negativere	negativere	airline	airline_ser	name	negativere	retweet_c	text	tweet_coord	tweet_cre	tweet_loc	user_timezone
2	5.7E+17	neutral	1			Virgin America		cairdin		0	@VirginAmerica What @dhepbu		########		Eastern Time (US & Canada)
3	5.7E+17	positive	0.3486		0	Virgin America		jnardino		0	@VirginAmerica plus you've add		########		Pacific Time (US & Canada)
4	5.7E+17	neutral	0.6837			Virgin America		yvonnalynn		0	@VirginAmerica I didn't today...		########	Lets Play	Central Time (US & Canada)
5	5.7E+17	negative	1	Bad Flight	0.7033	Virgin America		jnardino		0	@VirginAmerica it's really aggre		########		Pacific Time (US & Canada)
6	5.7E+17	negative	1	Can't Tell	1	Virgin America		jnardino		0	@VirginAmerica and it's a really		########		Pacific Time (US & Canada)
7	5.7E+17	negative	1	Can't Tell	0.6842	Virgin America		jnardino		0	@VirginA		########		Pacific Time (US & Canada)
8	5.7E+17	positive	0.6745		0	Virgin America		cjmcginnis		0	@VirginAmerica yes, nearly ever		########	San Franci	Pacific Time (US & Canada)
9	5.7E+17	neutral	0.634			Virgin America		pilot		0	@VirginAmerica Really missed a		########	Los Angele	Pacific Time (US & Canada)
10	5.7E+17	positive	0.6559			Virgin America		dhepburn		0	@virginamerica Well, I didn'tâ€¦		########	San Diego	Pacific Time (US & Canada)
11	5.7E+17	positive	1			Virgin America		YupitsTate		0	@VirginAmerica it was amazing,		########	Los Angele	Eastern Time (US & Canada)
12	5.7E+17	neutral	0.6769		0	Virgin America		idk_but_youtube		0	@VirginAmerica did you know th		########	1/1 loner s	Eastern Time (US & Canada)
13	5.7E+17	positive	1			Virgin America		HyperCamiLax		0	@VirginAmerica I <3 pretty gr		########	NYC	America/New_York
14	5.7E+17	positive	1			Virgin America		HyperCamiLax		0	@VirginAmerica This is such a gr		########	NYC	America/New_York
15	5.7E+17	positive	0.6451			Virgin America		mollanderson		0	@VirginAmerica @virginmedia I'		########		Eastern Time (US & Canada)
16	5.7E+17	positive	1			Virgin America		sjespers		0	@VirginAmerica Thanks!		########	San Franci	Pacific Time (US & Canada)
17	5.7E+17	negative	0.6842	Late Flight	0.3684	Virgin America		smartwatermelon		0	@VirginAmerica SFO-PDX sched		########	palo alto,	Pacific Time (US & Canada)
18	5.7E+17	positive	1			Virgin America		ItzBrianHunty		0	@VirginAmerica So excited for n		########	west covin	Pacific Time (US & Canada)
19	5.7E+17	negative	1	Bad Flight	1	Virgin America		heatherovieda		0	@VirginAmerica I flew from NY		########	this place	Eastern Time (US & Canada)
20	5.7E+17	positive	1			Virgin America		thebrandiray		0	I â¤ï, flying @VirginAmerica. â˜		########	Somewher	Atlantic Time (Canada)
21	5.7E+17	positive	1			Virgin America		JNLpierce		0	@VirginAmerica you know what		########	Boston | V	Quito
22	5.7E+17	negative	0.6705	Can't Tell	0.3614	Virgin America		MISSGJ		0	@VirginAmerica why are your fi		########		
23	5.7E+17	positive	1			Virgin America		DT_Les		0	@VirginAn	[40.74804263, -73.9	########		
24	5.7E+17	positive	1			Virgin America		ElvinaBeck		0	@VirginAmerica I love the hipste		########	Los Angele	Pacific Time (US & Canada)
25	5.7E+17	neutral	1			Virgin America		rjlynch21086		0	@VirginAmerica will you be mak		########	Boston, M	Eastern Time (US & Canada)
26	5.7E+17	negative	1	Customer	0.3557	Virgin America		ayeevickiee		0	@VirginAmerica you guys messe		########	714	Mountain Time (US & Canada)
27	5.7E+17	negative	1	Customer	1	Virgin America		Leora13		0	@VirginAmerica status match pr		########		
28	5.7E+17	negative	1	Can't Tell	0.6614	Virgin America		meredithjlynn		0	@VirginAmerica What happened		########		
29	5.7E+17	neutral	0.6854			Virgin America		AdamSinger		0	@VirginAmerica do you miss me		########	San Franci	Central Time (US & Canada)
30	5.7E+17	negative	1	Bad Flight	1	Virgin America		blackjackpro911		0	@VirginAn	[42.361016, -71.020	########	San Mateo, CA & Las Vegas, NV	
31	5.7E+17	neutral	0.615		0	Virgin America		TenantsUpstairs		0	@VirginAn	[33.94540417, -118.4	########	Brooklyn	Atlantic Time (Canada)



- The independent variables of sentiment analysis are categorized.

TECHNIQUES FOR ANALYSIS:

BAG OF WORD[BoW]:

- The value in each dimension represents the frequency of the word's occurrence in the document.
- To perform sentiment analysis, you can use these word frequency vectors to train a machine learning model (e.g., Naive Bayes, Logistic Regression) to classify text into positive, negative, or neutral sentiments.

WORD EMBEDDINGS:

- Word embeddings capture semantic relationships between words. Words with similar meanings are represented as vectors close to each other in this vector space.
- In sentiment analysis, the document is represented as a combination of the word vectors within it.
- Then, the document representations are used as features to train machine learning models for sentiment classification.

Transformer Models: BERT (Bidirectional Encoder Representations from Transformers):

- BERT (Bidirectional Encoder Representations from Transformers) takes into account the context of a word by considering both the left and right context[captures the meaning of words in a more sophisticated manner].
- Transformers are known for their state-of-the-art performance in NLP tasks and have the ability to capture nuanced sentiment, context, and sarcasm in text.

Traditional Method for Sentiment Analysis:

- Using reference dictionaries , the traditional sentiment analysis assigns a score for each of the words in a sentence. The average of these scores is calculated and the sentiment of the text is then interpreted.

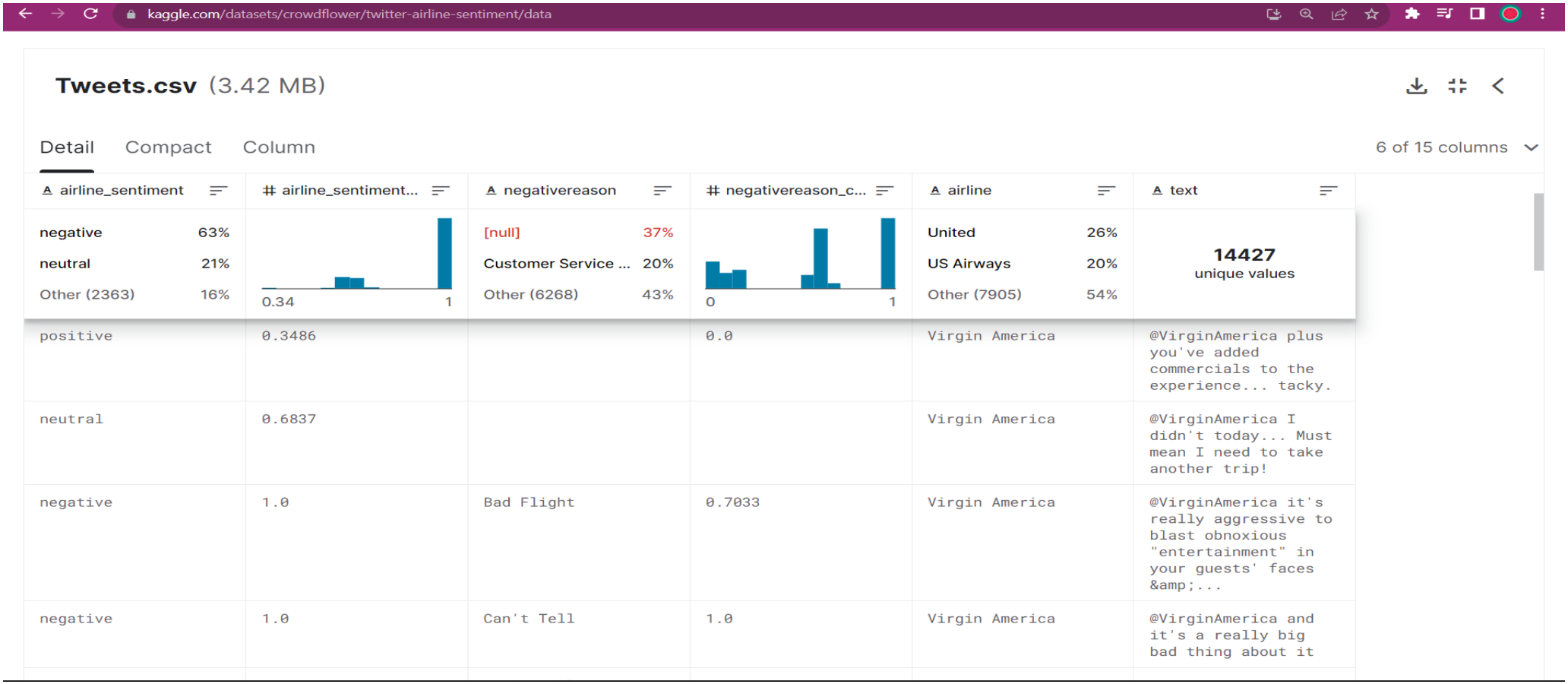
DEPLOYMENT OF MACHINE LEARNING:

- ML Ops is the process of deploying a machine learning model and integrating it into software that can be used by end users
- Machine learning involves giving data to a model to make predictions using statistical models, while deep learning uses artificial neural networks to process unstructured data like images, videos, and text.

- The common machine learning models:
    1. Logistic Regression
    2. Decision Trees
    3. Random Forest
    4. Support Vector Machines (SVM)
    5. k-NN
    6. Naive Bayes
    7. Neural Networks (Deep Learning)
    8. Clustering Algorithms
    9. Time Series Models

FEATURE EXTRACTION:

- We perform sentiment analysis using the independent variables and above ML models
- After we imputate the datset, the factors important for the sentiment in a response are filtered out.
- Consider, the tweet_id column in the dataset does not provide any form of sentiment. Likewise the values are noted down based on their weightage in the analysis.
- The dataset contains different reasons and texts, but the sentiments and sentiment confidence are groupable



VISUALIZATION AND INSIGHTS:

- After we analyze the dataset and create ML models, the data is visualized if necessary and the insights are generated/provided with reference to the data.

- In this project, we will use Artificial Neural Networks(ANN) model for the analysis of the dataset.

- After the successful development of the ML model, insights and visalization are obtained.

- The plotting of the dataset that has been analyzed is carried out using
    1. Pandas Plotting
    2. Matplotlib
    3. Seaborn