Sentiment Analysis

Lay Acheadeth & Misa Xirinda

November 16 2021

Existing problem:

Every business needs user feedback and review in order to improve their product to fulfill customer demands. Feedback has always been the key factor that drives business. From feedback, we can know the reaction of our customer to the product whether they are happy with it or not or if some revision needs to be made to the product.

In this project, we are going to explore the dataset and then find out the sentiment our customer has toward the product which is amazon baby product for this project.

Dataset:

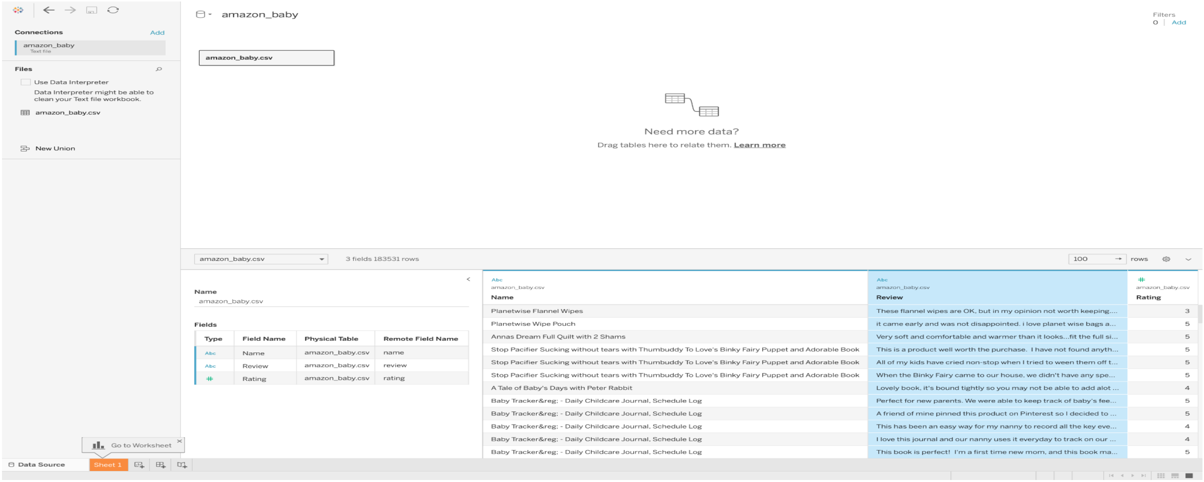
Our dataset is from kaggle. It's a product review on amazon baby products we obtain from kaggle.

Data preprocessing steps:

The data has only 3 columns: Name of the product, Review, and Rating.
The data itself cannot be dropped further since all the columns consist of important information.

Review column will be the input since we analyze those reviews to gain insight of user satisfaction with our amazon baby products. Rating will be used to determine the sentiment review. For example, If the review is likely positive, how much of it? Is it really positive? Or is it not?

1. The dataset has about 183,532 data.



3 fields 183531 rows

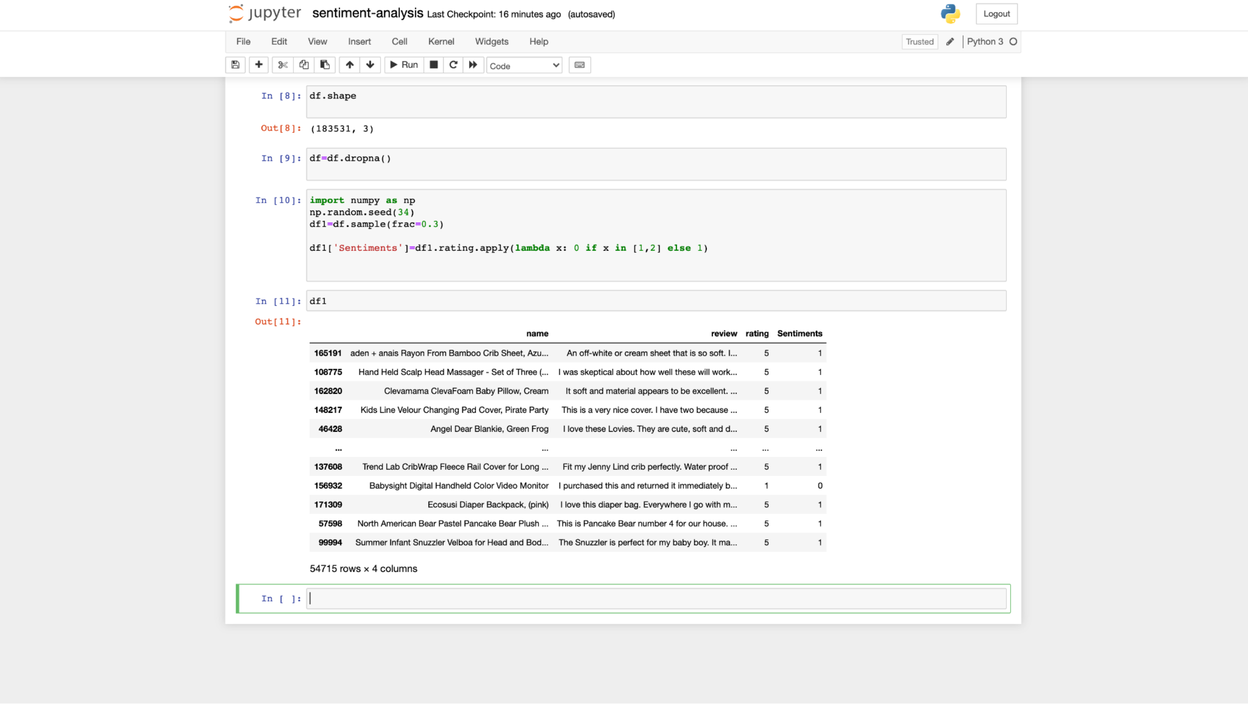
Name
amazon_baby.csv

Fields

Type	Field Name	Physical Table	Remote Field Name
Abc	Name	amazon_baby.csv	name
Abc	Review	amazon_baby.csv	review
#	Rating	amazon_baby.csv	rating

Abc amazon_baby.csv Name	Abc amazon_baby.csv Review	# amazon_baby.csv Rating
Planetwise Flannel Wipes	These flannel wipes are OK, but in my opinion not worth keeping....	3
Planetwise Wipe Pouch	it came early and was not disappointed. i love planet wise bags a...	5
Annas Dream Full Quilt with 2 Shams	Very soft and comfortable and warmer than it looks...fit the full si...	5
Stop Pacifier Sucking without tears with Thumbuddy To Love's Binky Fairy Puppet and Adorable Book	This is a product well worth the purchase. I have not found anyth...	5
Stop Pacifier Sucking without tears with Thumbuddy To Love's Binky Fairy Puppet and Adorable Book	All of my kids have cried non-stop when I tried to ween them off t...	5
Stop Pacifier Sucking without tears with Thumbuddy To Love's Binky Fairy Puppet and Adorable Book	When the Binky Fairy came to our house, we didn't have any spe...	5
A Tale of Baby's Days with Peter Rabbit	Lovely book, it's bound tightly so you may not be able to add alot ...	4
Baby Tracker&reg; - Daily Childcare Journal, Schedule Log	Perfect for new parents. We were able to keep track of baby's fee...	5
Baby Tracker&reg; - Daily Childcare Journal, Schedule Log	A friend of mine pinned this product on Pinterest so I decided to ...	5
Baby Tracker&reg; - Daily Childcare Journal, Schedule Log	This has been an easy way for my nanny to record all the key eve...	4
Baby Tracker&reg; - Daily Childcare Journal, Schedule Log	I love this journal and our nanny uses it everyday to track on our ...	4
Baby Tracker&reg; - Daily Childcare Journal, Schedule Log	This book is perfect! I'm a first time new mom, and this book ma...	5

2. This dataset is pretty big. Hence, we will only use 30% of the overall data since working with all the data will be time consuming and does not guarantee better performance.



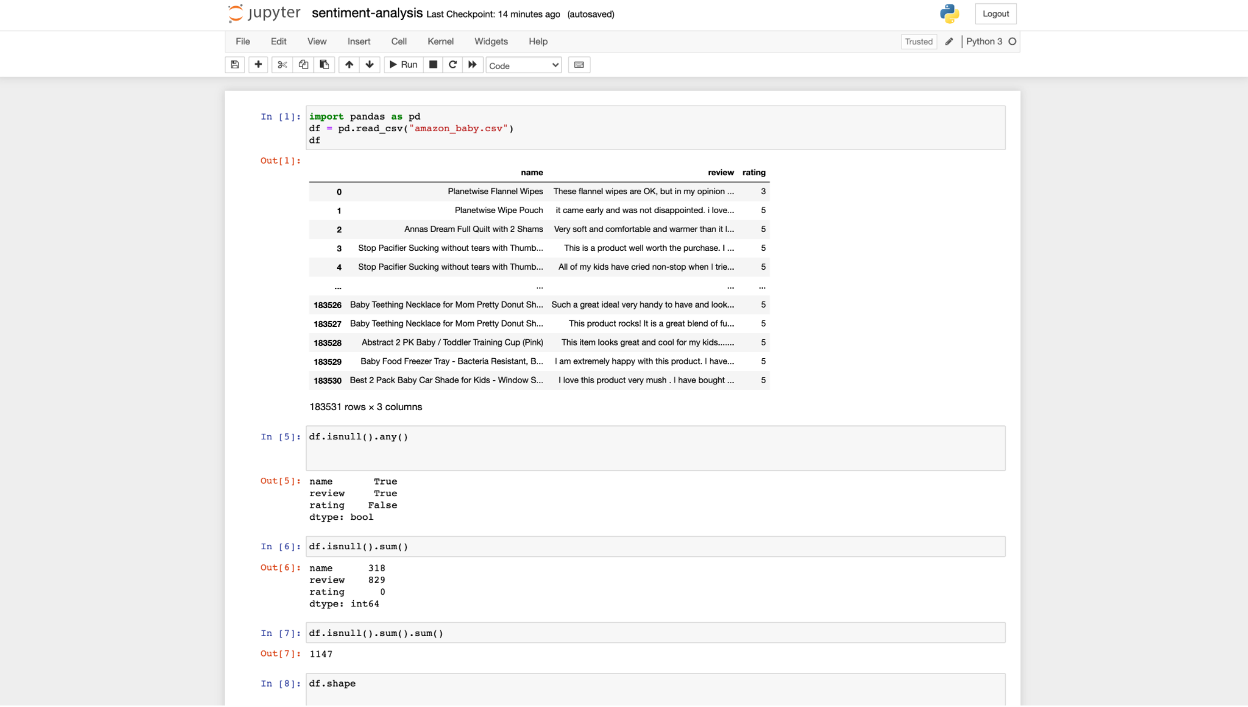
```
In [8]: df.shape

Out[8]: (183531, 3)

In [9]: df=df.dropna()

In [10]: import numpy as np
         np.random.seed(34)
         df1=df.sample(frac=0.3)

         df1['Sentiments']=df1.rating.apply(lambda x: 0 if x in [1,2] else 1)

In [11]: df1
Out[11]:
```

	name	review	rating	Sentiments
165191	aden + anais Rayon From Bamboo Crib Sheet, Azu...	An off-white or cream sheet that is so soft. I...	5	1
108775	Hand Held Scalp Head Massager - Set of Three (...	I was skeptical about how well these will work...	5	1
162820	Clevamama ClevaFoam Baby Pillow, Cream	It soft and material appears to be excellent. ...	5	1
148217	Kids Line Velour Changing Pad Cover, Pirate Party	This is a very nice cover. I have two because ...	5	1
46428	Angel Dear Blankie, Green Frog	I love these Lovies. They are cute, soft and d...	5	1
...	...	...	...	...
137608	Trend Lab CribWrap Fleece Rail Cover for Long ...	Fit my Jenny Lind crib perfectly. Water proof ...	5	1
156932	Babysight Digital Handheld Color Video Monitor	I purchased this and returned it immediately b...	1	0
171309	Ecosusi Diaper Backpack, (pink)	I love this diaper bag. Everywhere I go with m...	5	1
57598	North American Bear Pastel Pancake Bear Plush ...	This is Pancake Bear number 4 for our house. ...	5	1
99994	Summer Infant Snuzzler Velboa for Head and Bod...	The Snuzzler is perfect for my baby boy. It ma...	5	1

54715 rows × 4 columns

```
In [ ]:
```

3. Our data has an overall of 1147 null value, with 318 in name column and 829 in review. We decided to drop this null value since it's a text value. We cannot replace it with means and using another method does not guarantee efficiency than to drop it.



```
In [1]: import pandas as pd
        df = pd.read_csv("amazon_baby.csv")
        df
Out[1]:
```

	name	review	rating
0	Planetwise Flannel Wipes	These flannel wipes are OK, but in my opinion ...	3
1	Planetwise Wipe Pouch	it came early and was not disappointed. i love...	5
2	Annas Dream Full Quilt with 2 Shams	Very soft and comfortable and warmer than it l...	5
3	Stop Pacifier Sucking without tears with Thumb...	This is a product well worth the purchase. I ...	5
4	Stop Pacifier Sucking without tears with Thumb...	All of my kids have cried non-stop when I trie...	5
...	...	...	...
183526	Baby Teething Necklace for Mom Pretty Donut Sh...	Such a great idea! very handy to have and look...	5
183527	Baby Teething Necklace for Mom Pretty Donut Sh...	This product rocks! It is a great blend of fu...	5
183528	Abstract 2 PK Baby / Toddler Training Cup (Pink)	This item looks great and cool for my kids.......	5
183529	Baby Food Freezer Tray - Bacteria Resistant, B...	I am extremely happy with this product. I have...	5
183530	Best 2 Pack Baby Car Shade for Kids - Window S...	I love this product very mush . I have bought ...	5

183531 rows × 3 columns

```
In [5]: df.isnull().any()

Out[5]: name       True
        review     True
        rating    False
        dtype: bool

In [6]: df.isnull().sum()
Out[6]: name      318
        review    829
        rating      0
        dtype: int64

In [7]: df.isnull().sum().sum()
Out[7]: 1147

In [8]: df.shape
```

Since we are working with text data, we will have to convert that text to numerical value since most machine learning algorithms we are going to work with only accept numeric values.

Therefore, I am going to use Countvectorizer to convert the review data to numerical value.

CountVectorizer:
I will use a count vectorizer to vectorize the text data in the review column(training feature for the project).

- FIrst step: split the data into training sets and testing sets.
- Second step: vectorize the input feature that is out review column (both training and testing data)
- Import model
- Find the accuracy score (accuracy, confusion matrix).
- Find the true positive and negative rate

TFID Vectorizer:
This is another vectorizer technique that I am going to use that is known to be more popular because it uses the term frequency of the words.

The procedure to conduct this vectorizer is going to be the same as the previous one. Only the vectorizer is different.

Model and techniques:

- Logistic Regression: The idea is to divide the training set into positive and negative comments. Count all the words and make a python dictionary of their frequencies in positive and negative comments.
- Support vector machine: Used for both classification and regression.
- Decision tree: