**Symbiosis Skills and Professional University Kiwale, Pune**

------------------------------------------------------------------------------------------

**PROJECT REPORT**

**On**

**“Travel And Transportation”**

****

**Submitted by**

**Manali Ramesh Gaikwad**

**(2001207035)**

**DA-1-(Data Associate)**

**Under The Guidance of**

**Trainer's Name: Mr. Kushal Sharma**

**STUDENT DECLARATION AND ATTESTATION BY TRAINER**

This is to declare that this report has been written by me. No part of the report is plagiarized from other sources. All information included from other sources have been duly acknowledged. I aver that if any part of the report is found to be plagiarized, I shall take full responsibility for it.

Signature of student

**Name of student: Gaikwad Manali Ramesh**

Registration Number: 2001207035

Signature of trainer

**Name of trainer: Mr. Kushal Sharma**

**CERTIFICATE**

This is to certify that the report entitled, “**Travel And Transportation**” submitted by **”Manali Ramesh Gaikwad**” to Symbiosis Skills and Professional University, Pune, Maharashtra, India, is a record of bonafide Project work carried out by him under my supervision and guidance and is worthy of consideration for the completion of certificate course in “Data Associate”.

Signature of Trainer

Name of Trainer: Mr.Kushal Sharama

Date: / / 2021

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Supervisor Supervisor

Date:

**ACKNOWLEDGEMENTS**

The project report is a result of a dedicated work aspiring to master the concepts of Data Analysis. We would like to thank the people for their valuable guidance, sharing their valuable knowledge and precious time.

We would especially like to thank my entire class members for helping in this project and bringing this to completion.

We would also like to extend my sincere thanks to my project guide **Mr.Kushal Sharama** for him excellence guidance, invaluable suggestions and encouragement at stages of project work.

We would like to express my heart-felt gratitude to all my teachers who with their invaluable teaching have amplified the level of knowledge and helped a lot to complete this project.

**Name of Student:**

Manali Ramesh Gaikwad

**INDEX**

|  |  |  |
| --- | --- | --- |
| **SR NO.** | **CONTENT** | **Page No.** |
| 1. | Introduction | 6 |
| 2. | Purpose of Project | 7 |
| 3. | Problem Statement | 8 |
| 4. | Objective | 9 |
| 5. | Data Collection | 10 |
| 6. | Result | 11 |
| 7. | Conclusion | 12 |
| 8. | Annexure | 13 |
|  | 8.1 Screenshots |  |

**Introduction**

In recent years, Twitter has become the de facto online customer service platform. Thus, a companys image on Twitter is of central importance and this is especially true for airlines given that many tweets are travel-related in nature. In fact, research has shown that responding to tweets has revenue generating potential, drives higher satisfaction than other customer service channels, and perhaps most importantly, satisfied Twitter users spread the word.

In this project, we use tweets gathered from Twitter to learn about people’s flight experiences and give airline companies suggestions on how to make their trip more enjoyable. The data set contains about 15,000 tweets, collected from February 2015 on various airline reviews. Every review is labeled as either positive, negative or neutral. First, we want to build a model to perform sentiment analysis on the data set. Second, more interestingly, we want to assign a reason to each negative response. In our data set, about 80% of the negative reviews has a negative reason label, yet the rests are labeled as ”can’t tell”.

We use different types of machine learning algorithms for analysis. We perform this in python and tableau,power BI,Pig,Hive.

**Purpose of Project**

In this project, we aim to understand customer complaints on social media, specifically Twitter, regarding the major airlines in the US.

We get Know where we are doing good and where we are doing bad.can be a basis for a predictive model when we associated tweets length with sentiment.

Companies can get to know their competition.Improve the flight journey overall.

**Problem Statement**

* Given tweets about six US airlines, the task is to predict whether a tweet contains positive, negative, or neutral sentiment about the airline.
* To calculate accuracy using Random Forest algorithms.

**Objective**

* To identify the positive ,negative and neutral Sentiments of the customers about the airlines.
* Which airlines gets the most tweets.
* To calculate accuracy using Random Forest algorithms.

**Data Collection**

Airline tweets:

**Columns:**

1. tweet\_id

2. airline\_sentiment

3. airline\_sentiment\_confidence

4. negativereason

5. negativereason\_confidence

6. airline

7. airline\_sentiment\_gold

8. name

9. negativereason\_gold

10. retweet\_count

11. text

12. tweet\_coord

13. tweet\_created

14. tweet\_location

15. use\_timezone

**Result**

1. Random forest Accurancy:0.75

2. predict whether a tweet contains positive, negative, or neutral sentiment about the airline.

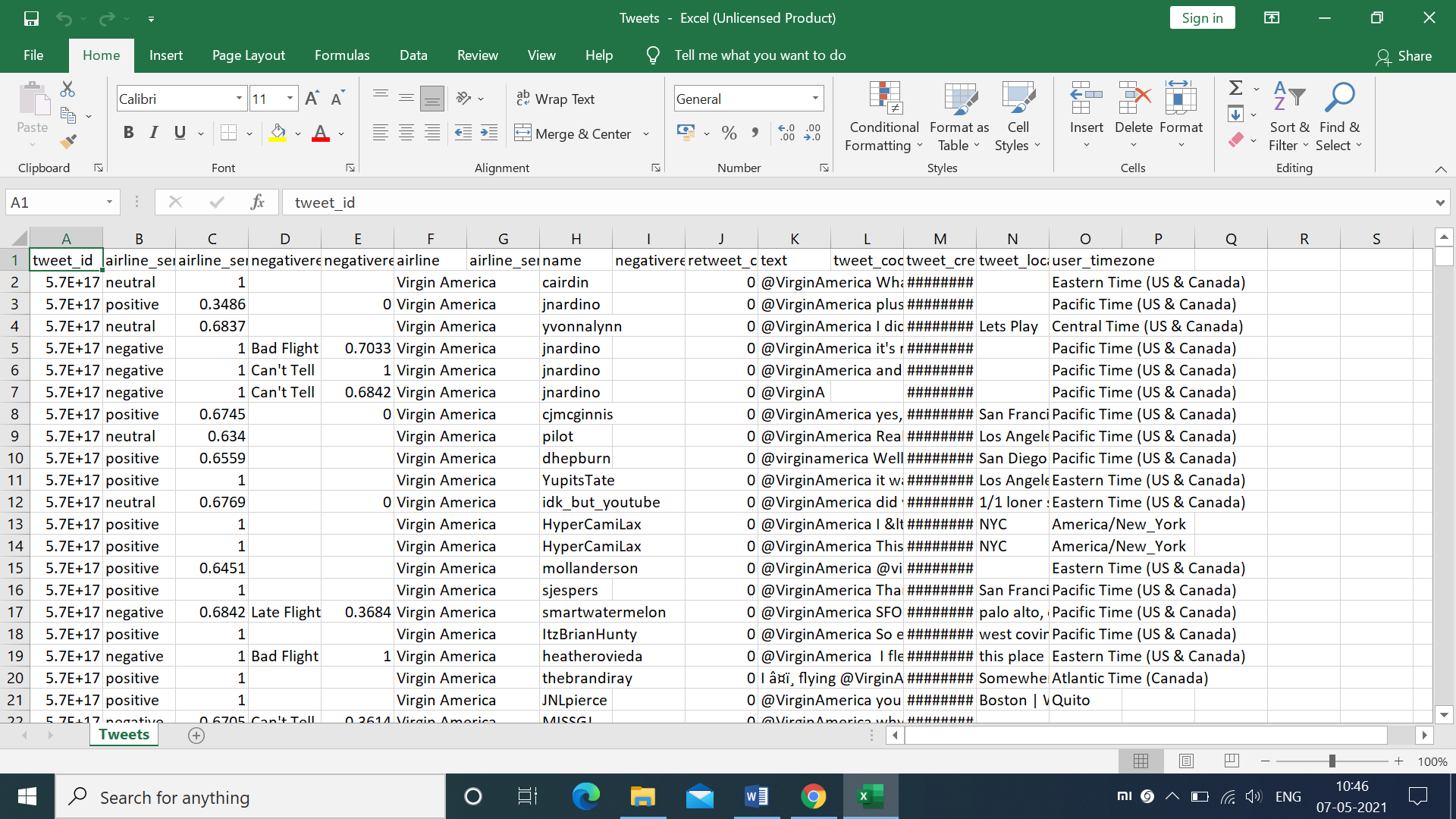
**Conclusion**

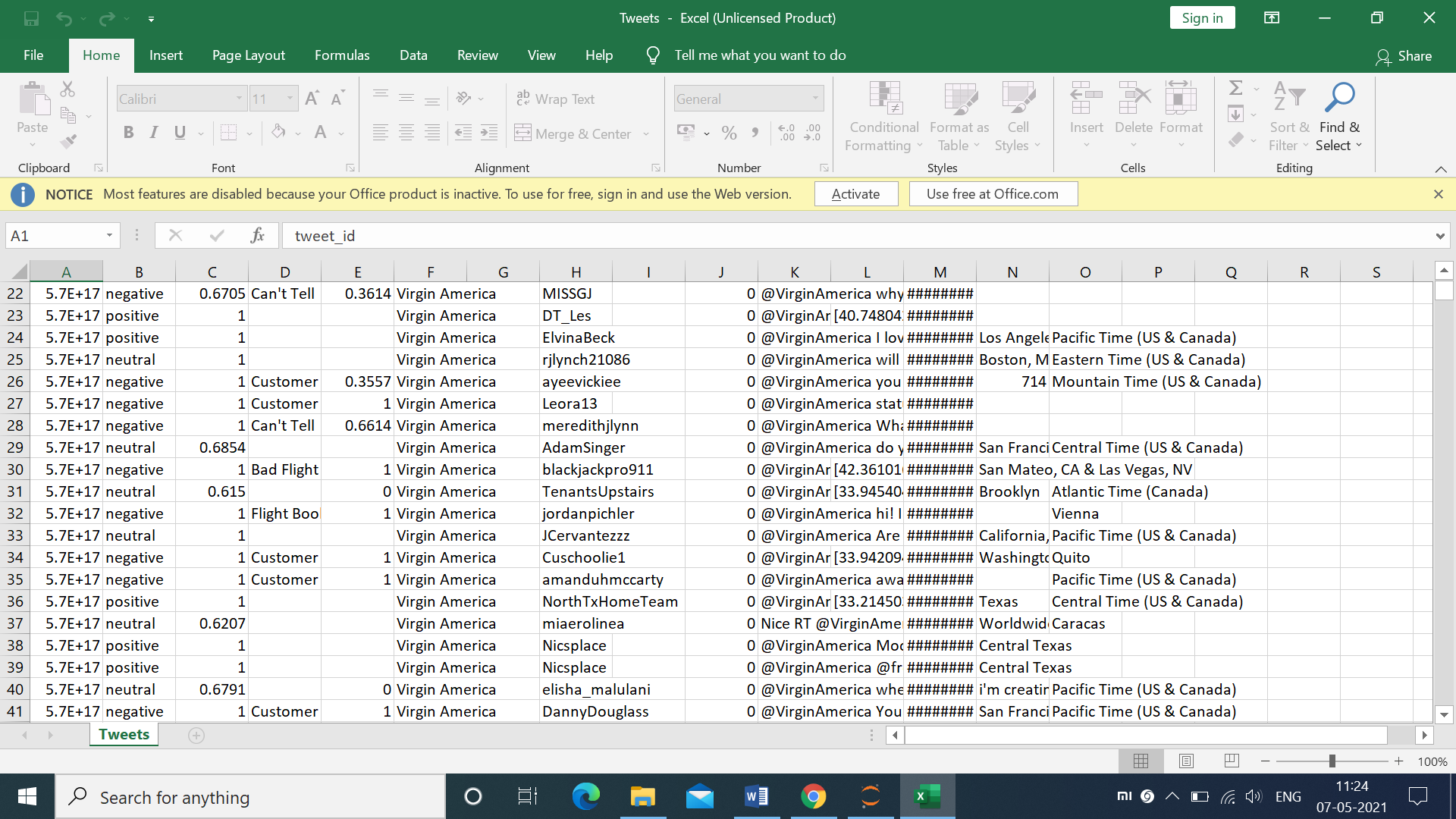
The task is to predict whether a tweet contains positive, negative, or neutral sentiment about the airline using python, tableau. For prediction numpy, pandas, matplotlib python library applied on that dataset.we will use machine learning algorithms to train and test our sentiment analysis models. We performed an analysis of public tweets regarding six US airlines and achieved an accuracy of around 75%.

**7.Annexure**

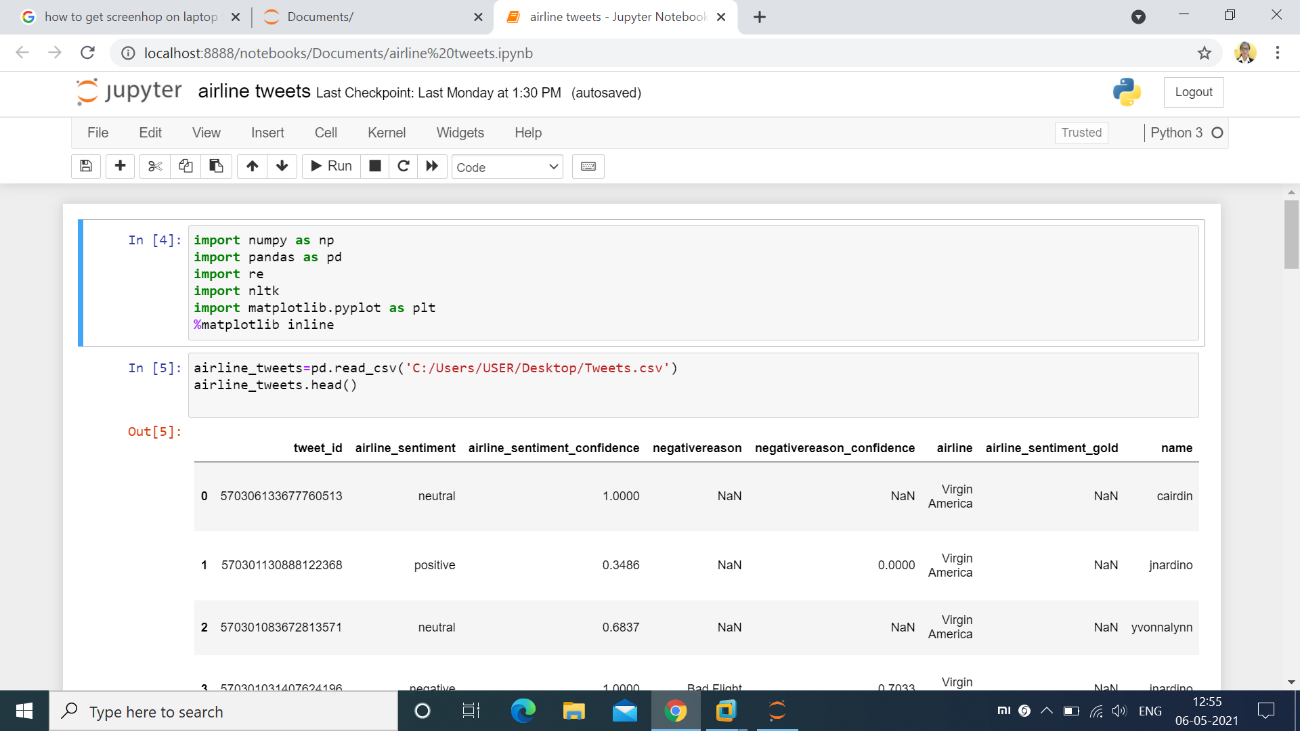
**7.1 Screenshots**

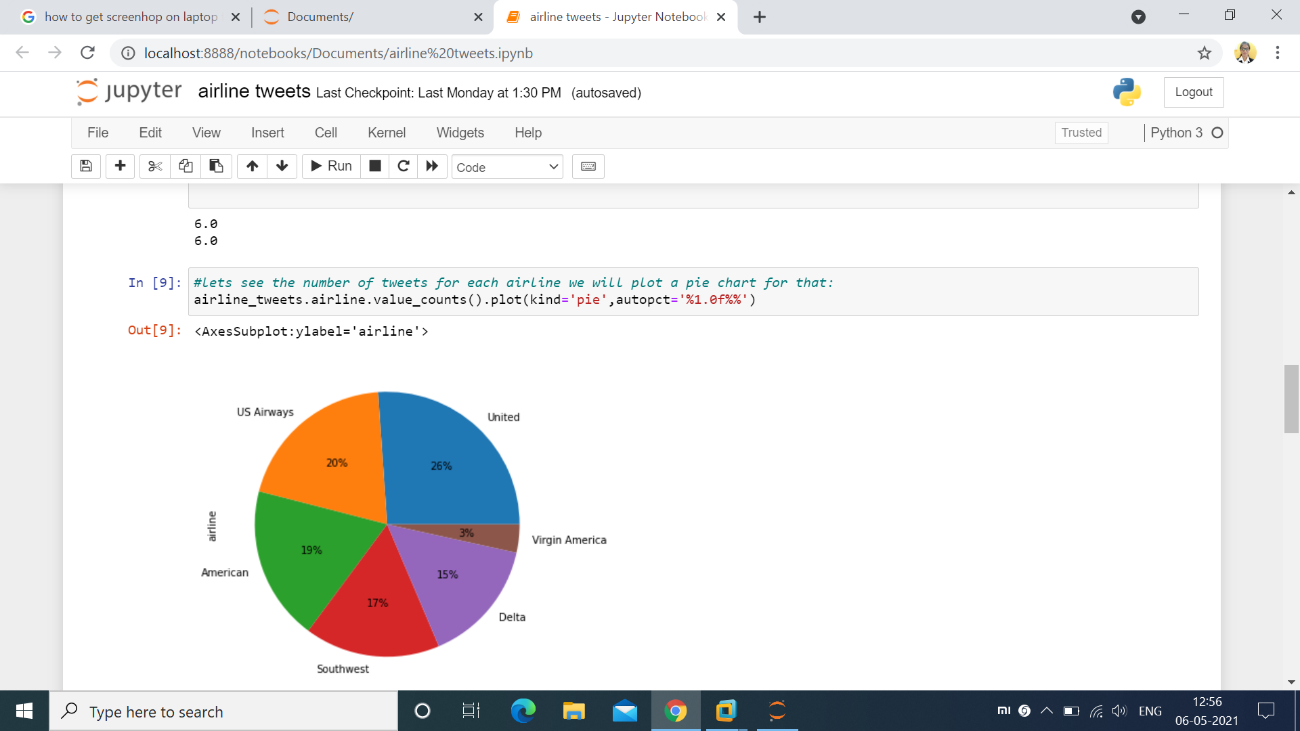
**1.Dataset**

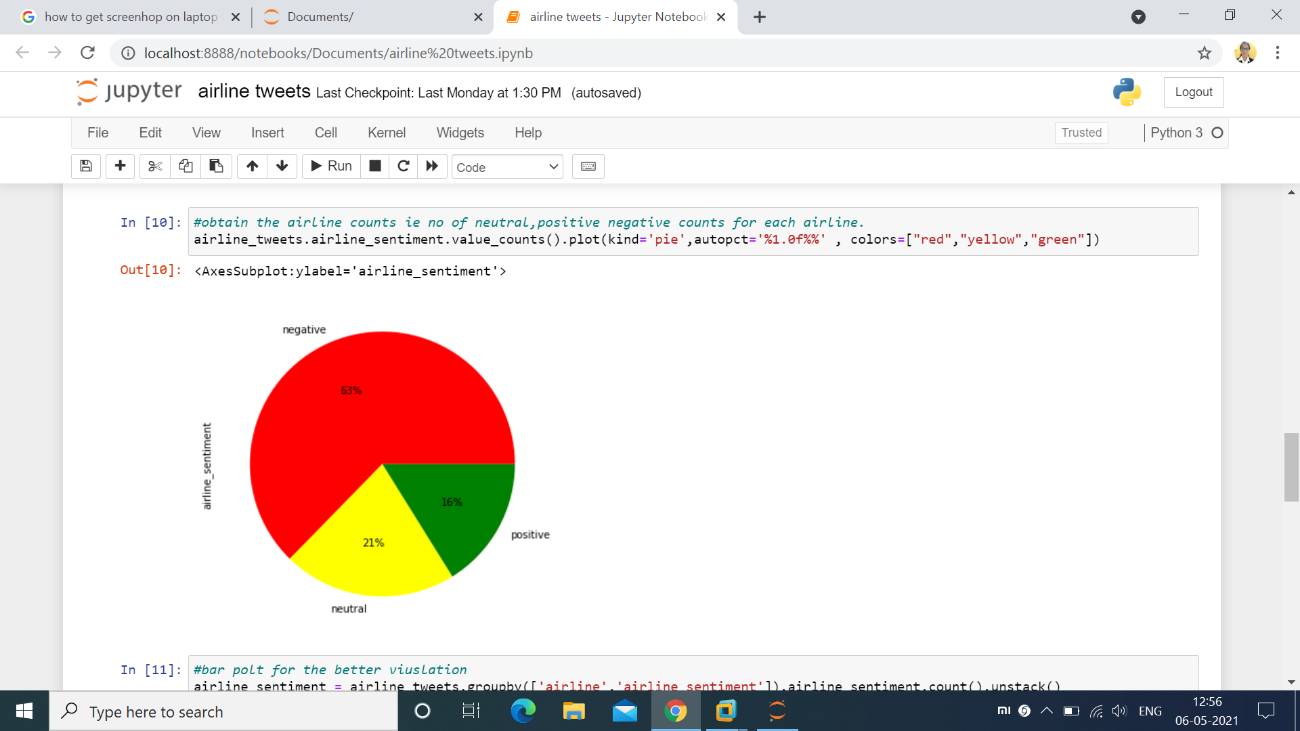
****

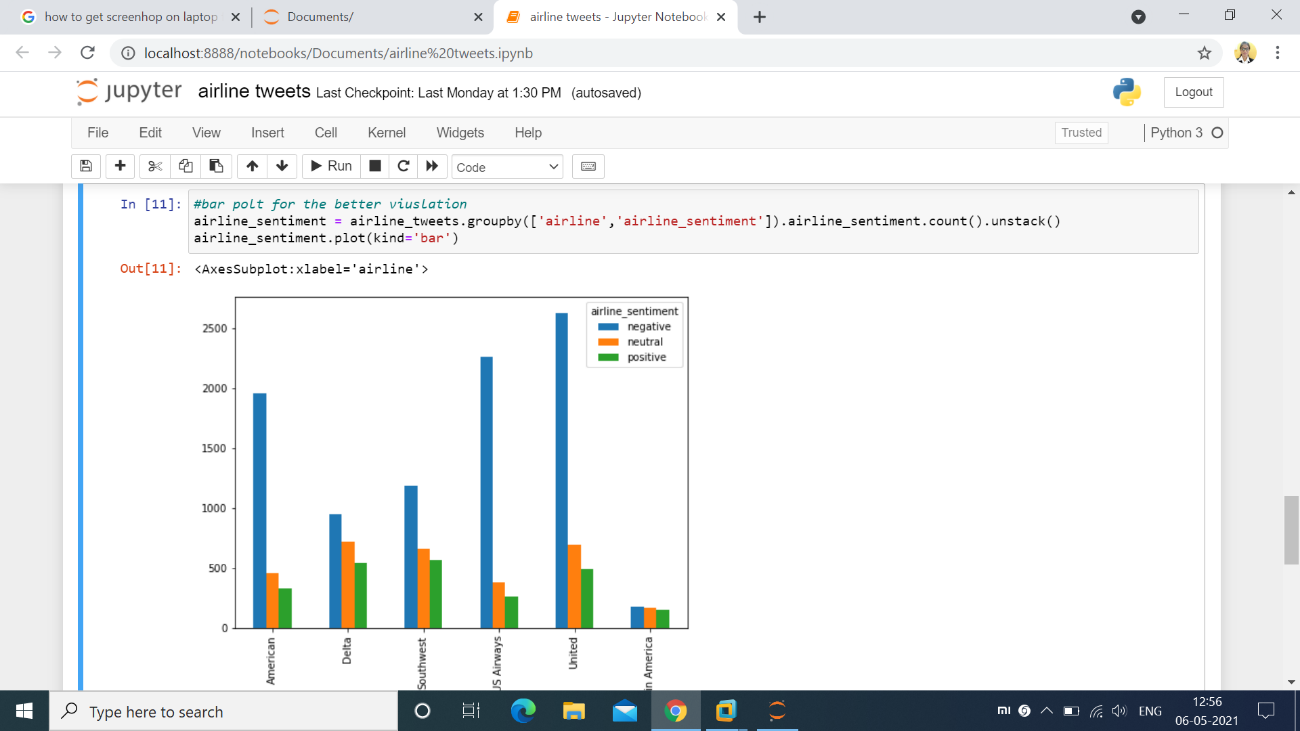
****

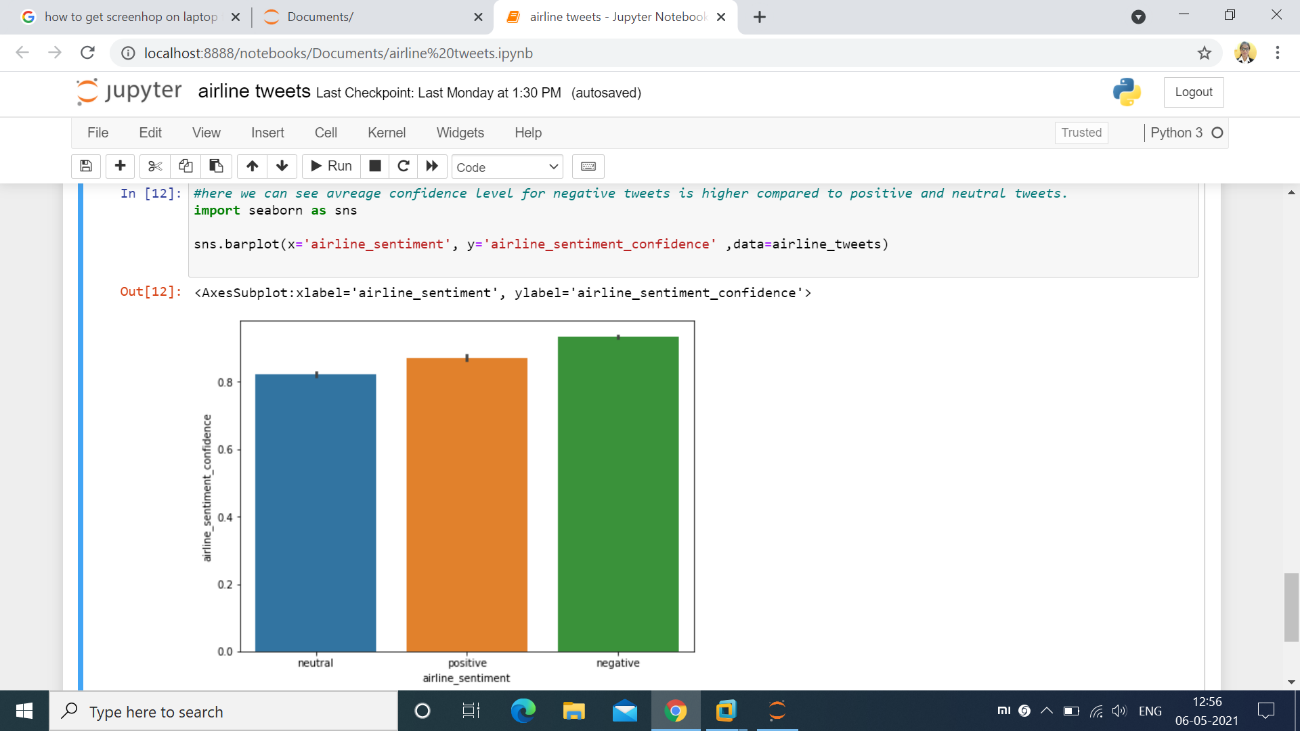
**2.Python**

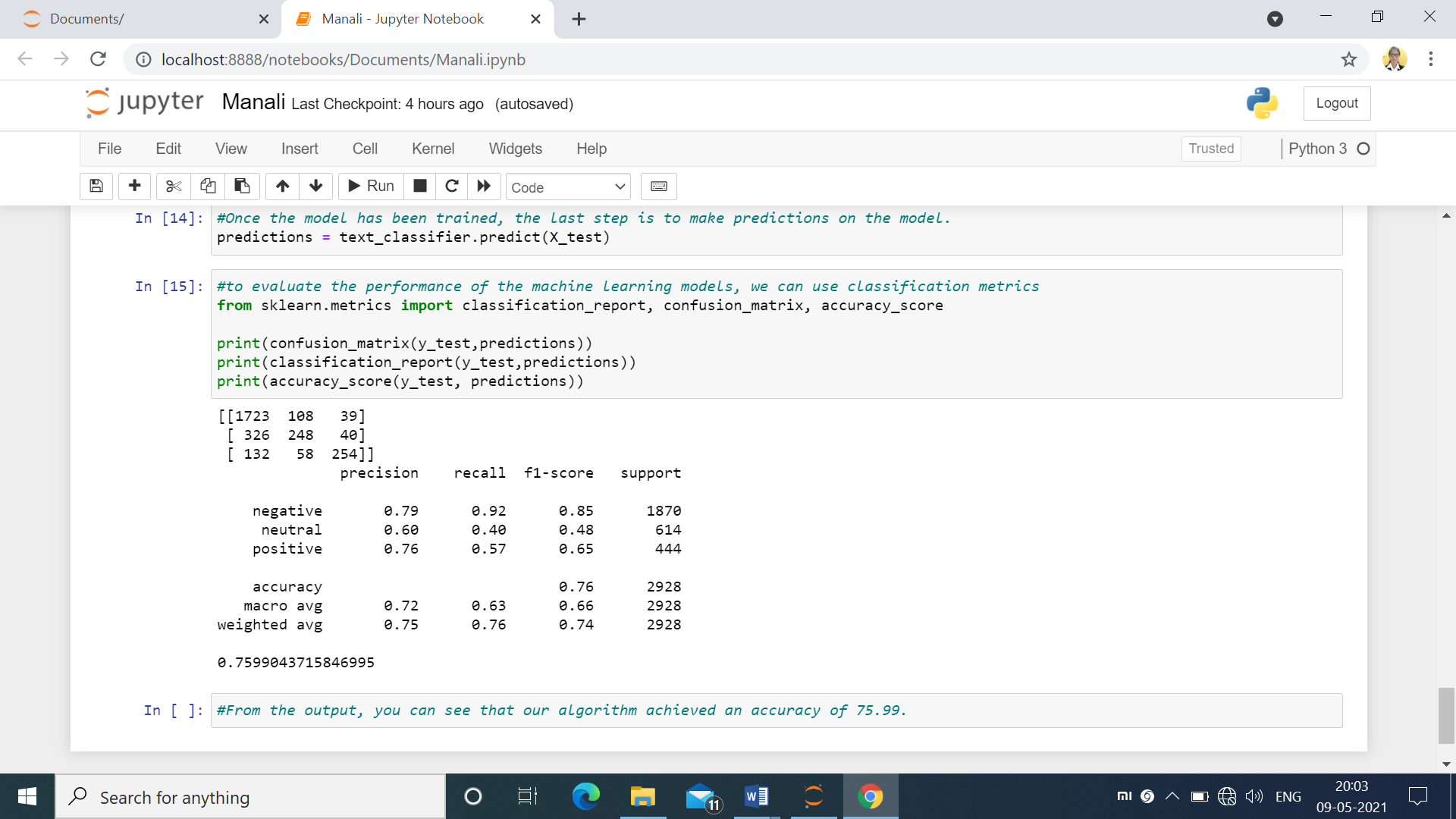
****

****

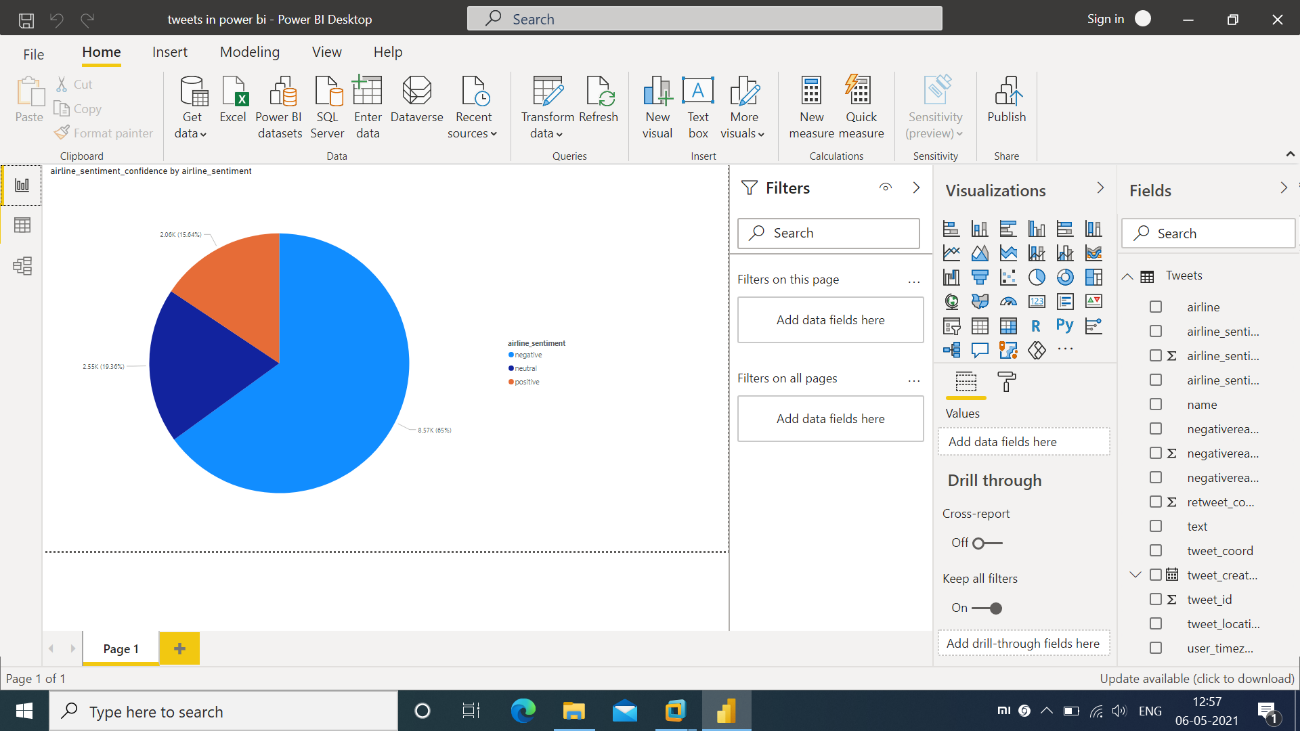
****

****

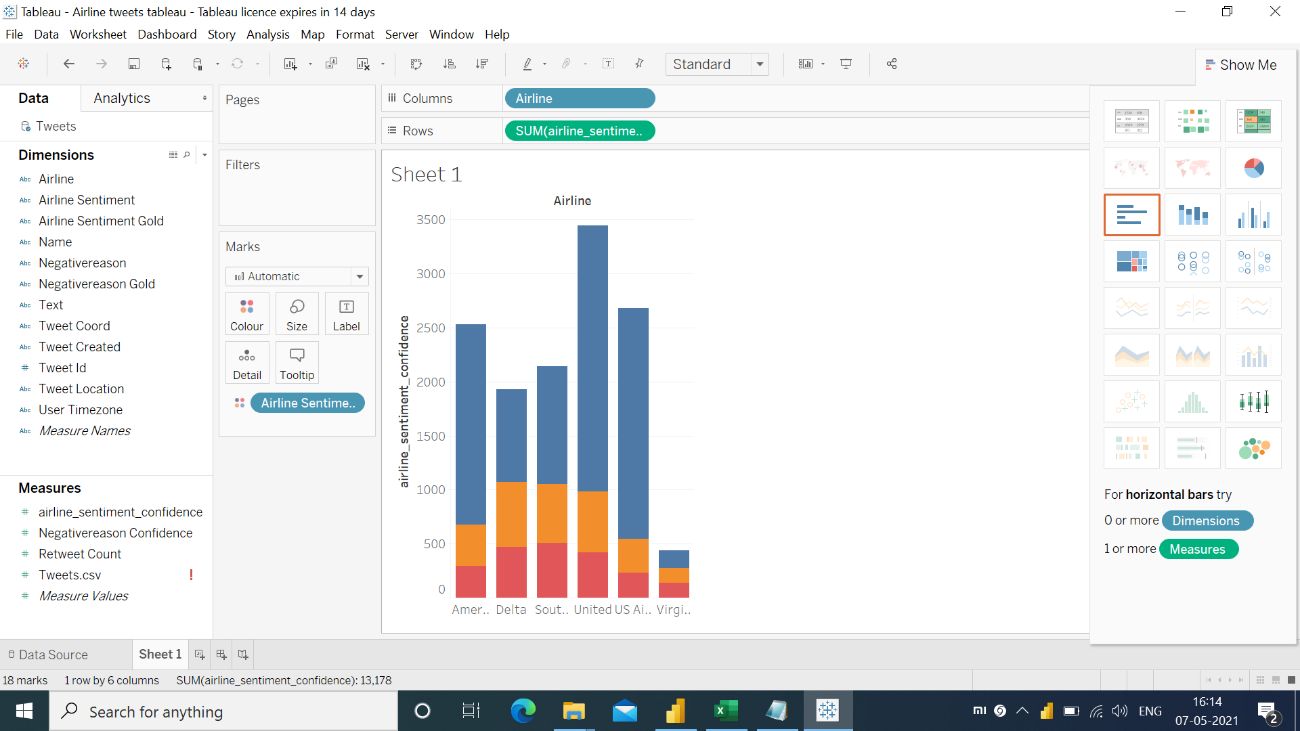
****

****

**3.Power BI**

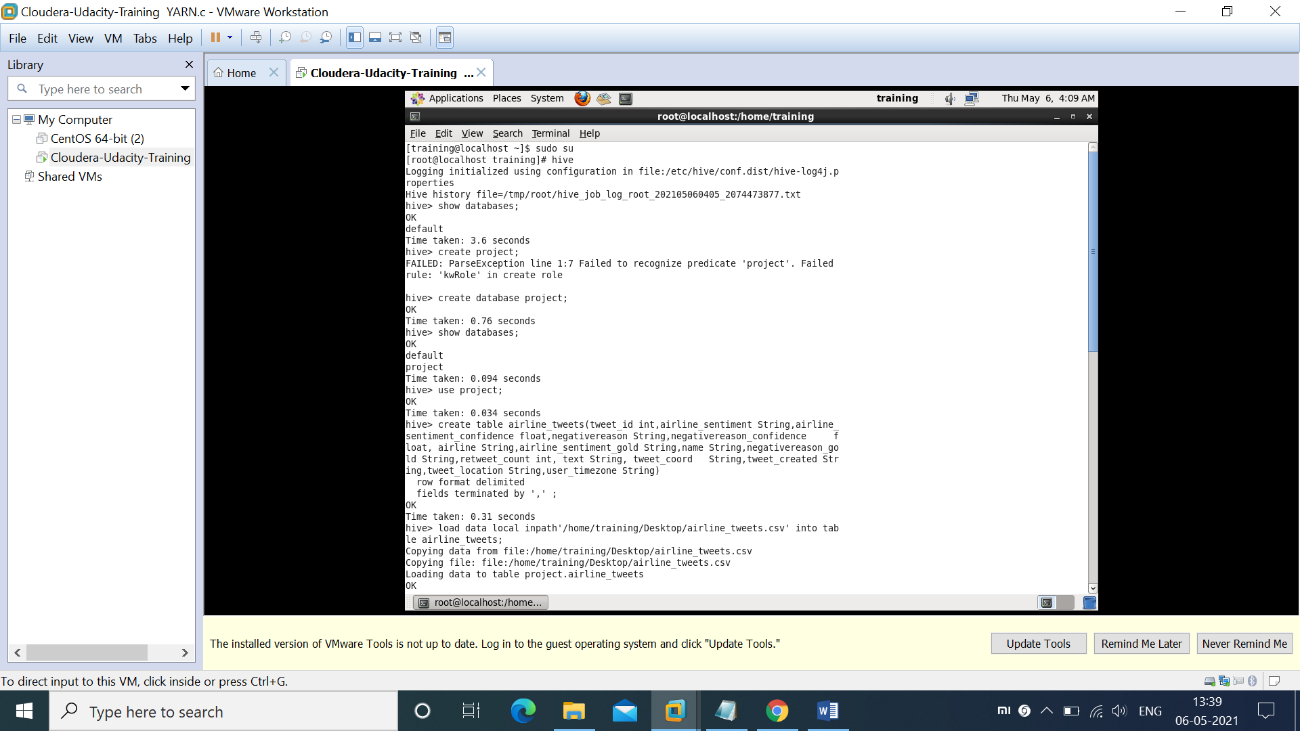
****

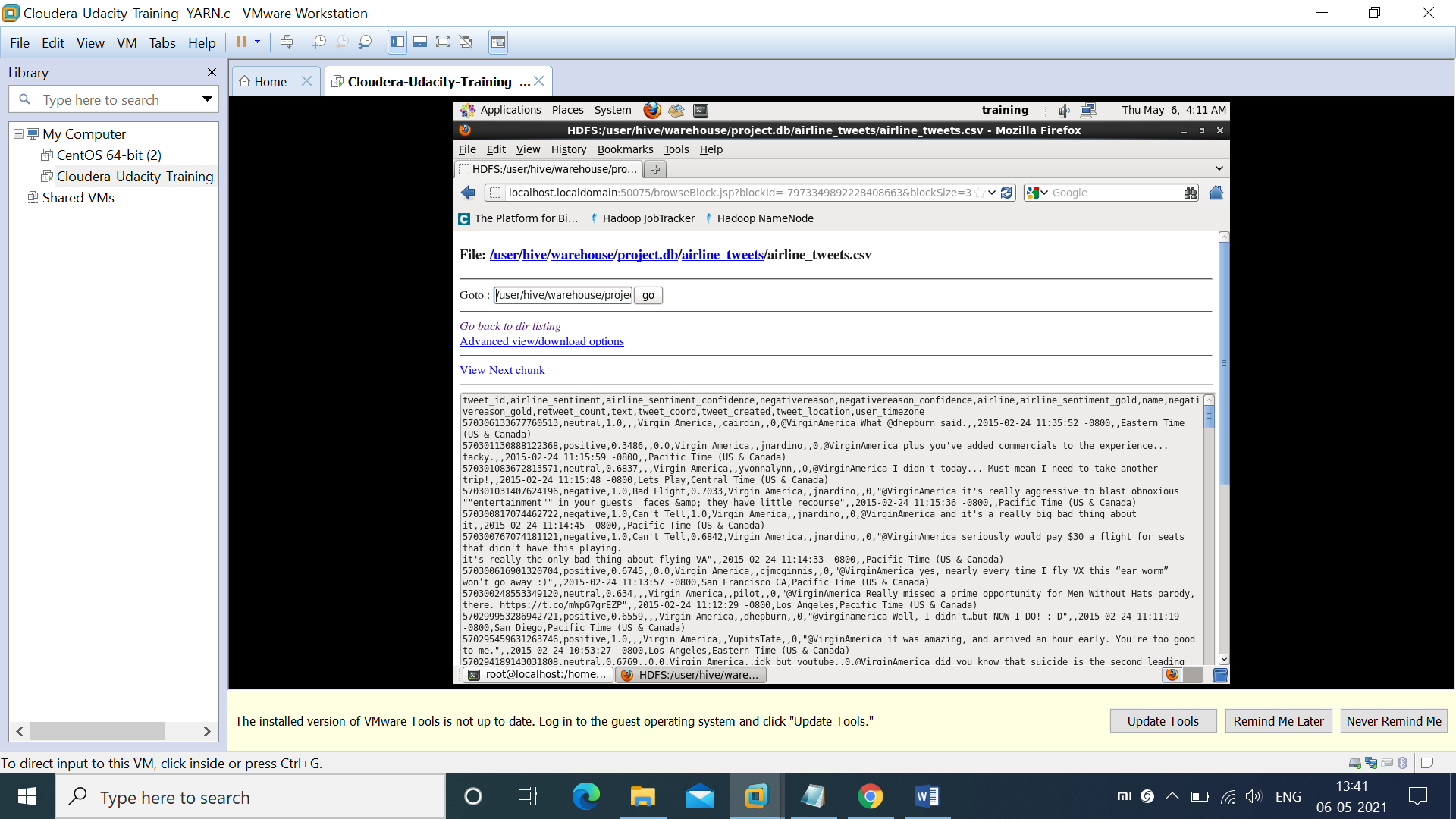
**4.Tableu**

****

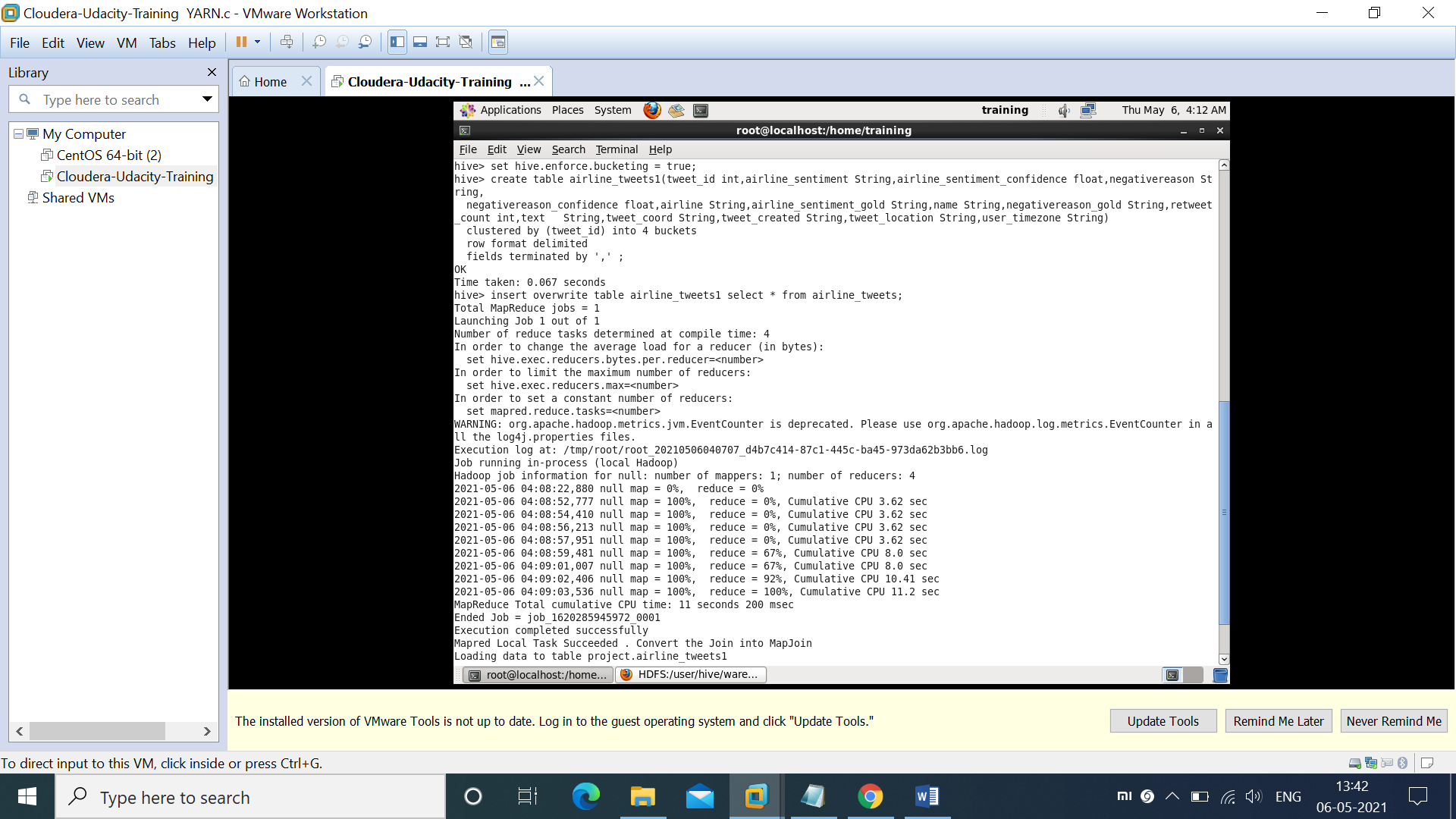
**5.Hive**

Creating table in hive and loading a data:-

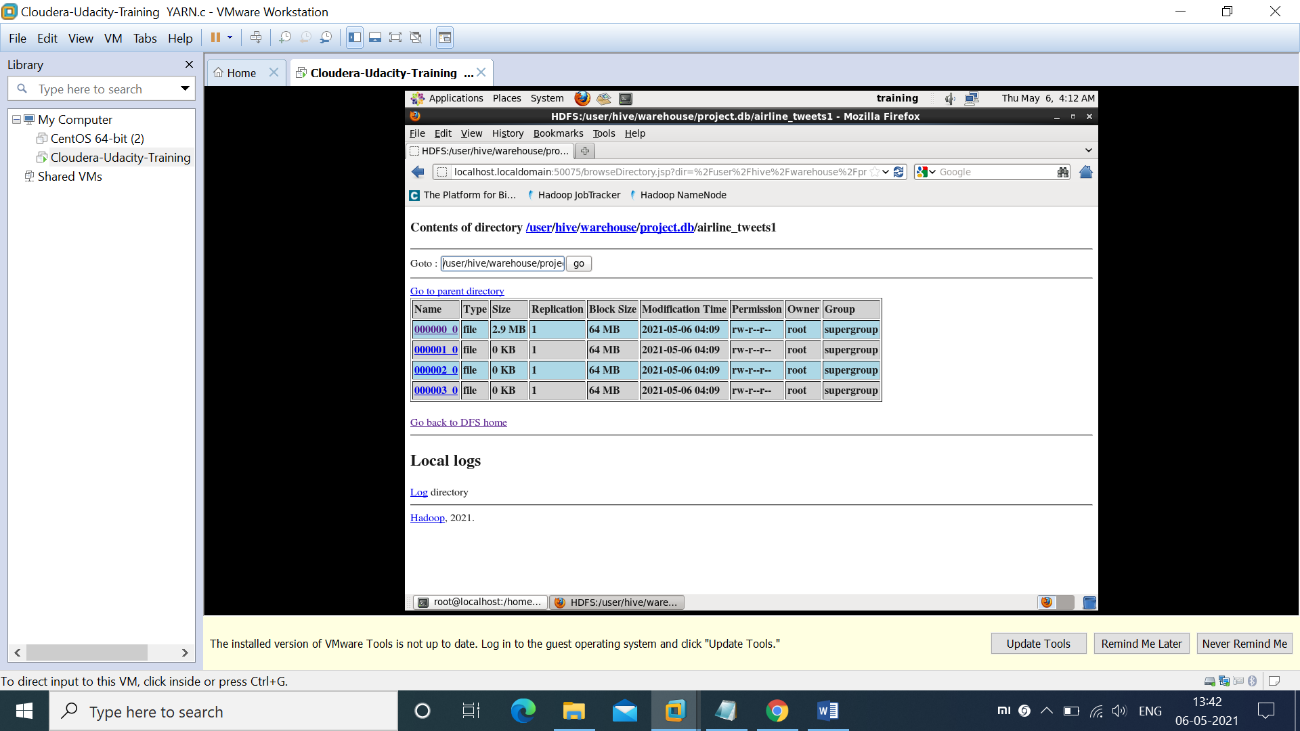




Bucketing example in hive:-

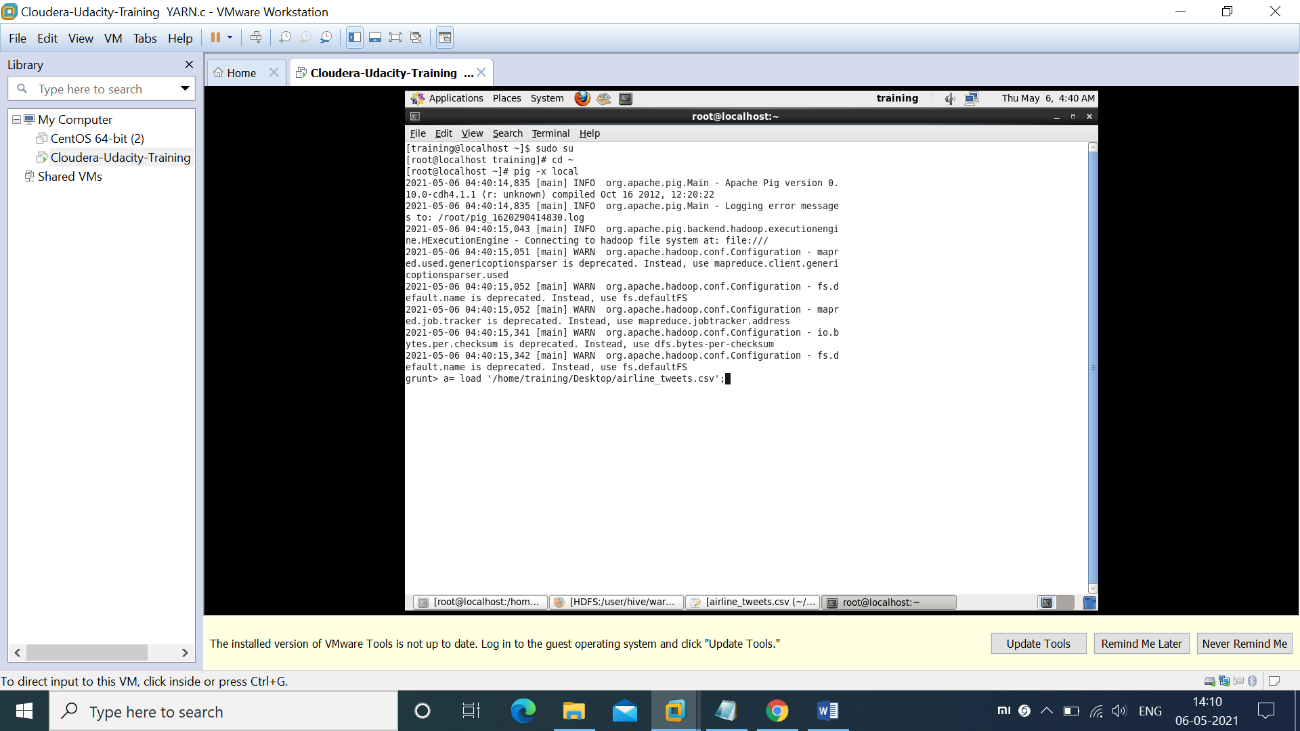


Final Output:-

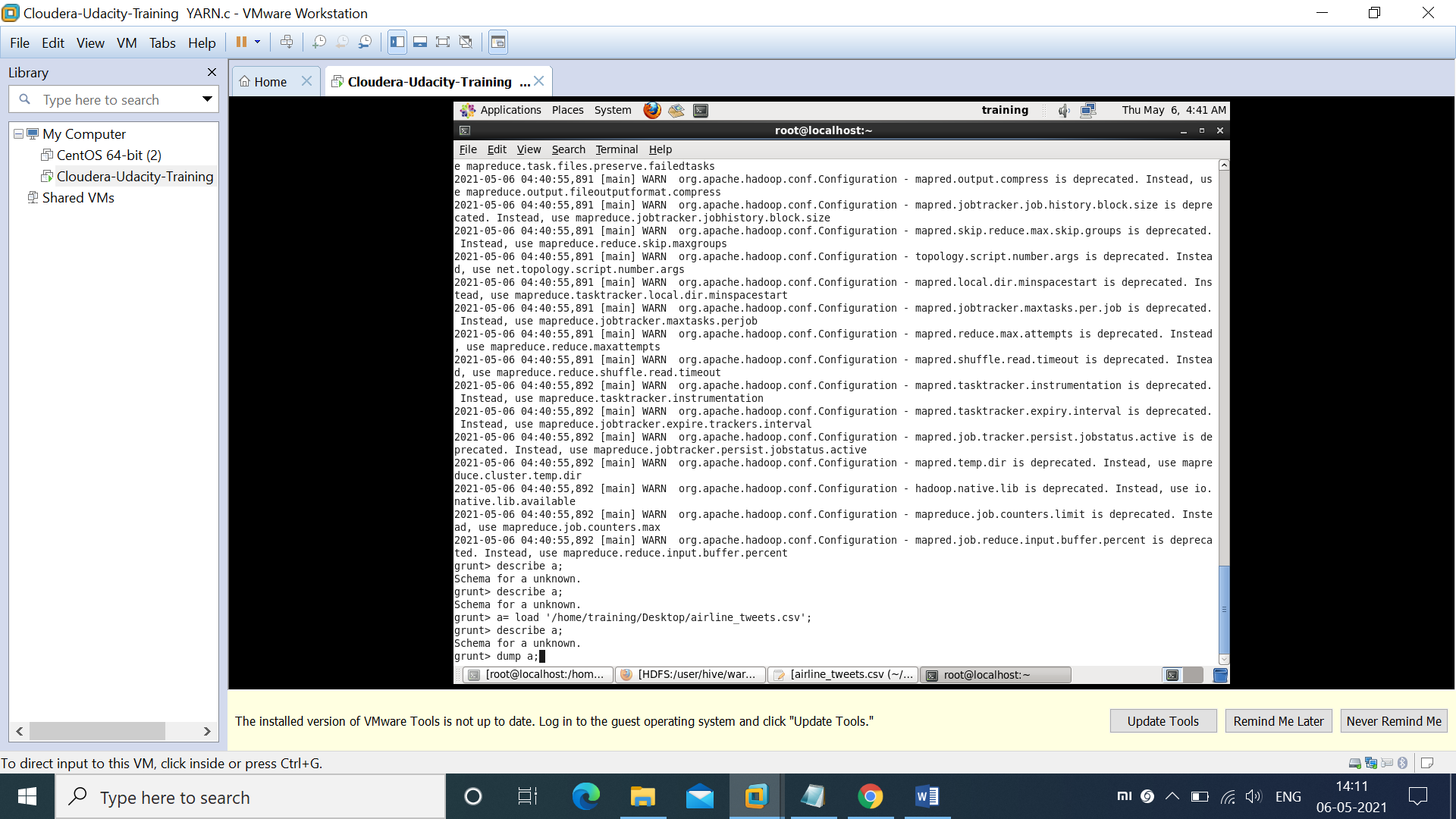
****

**6.Pig**

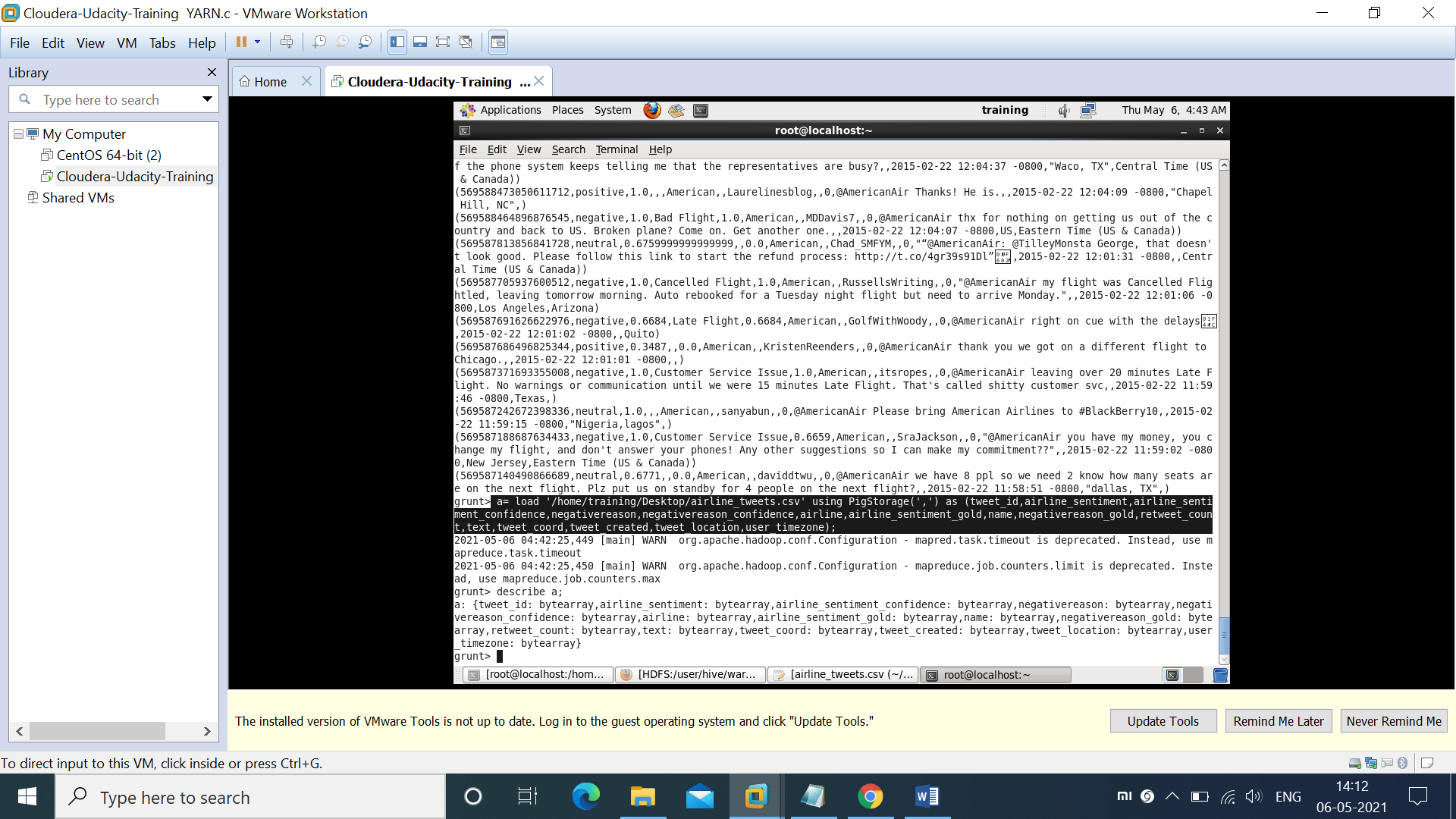
Loading a data into pig and creating table:-



Describing a data:



Defining schema without specifying any data type:-



Accessing the fields from dataset:-

