System Requirements Specification Index

For

Airline sentiment analysis Usecase

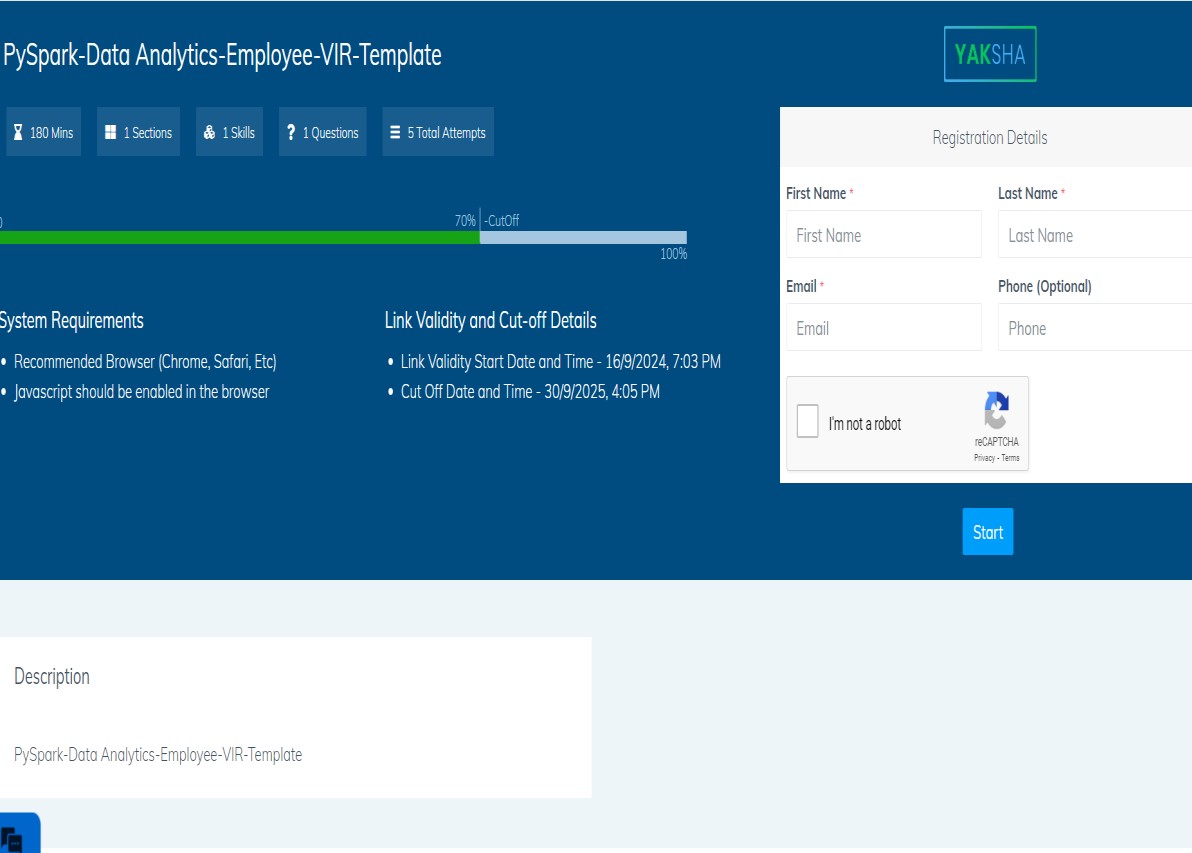
1.0

IIHT Pvt. Ltd.

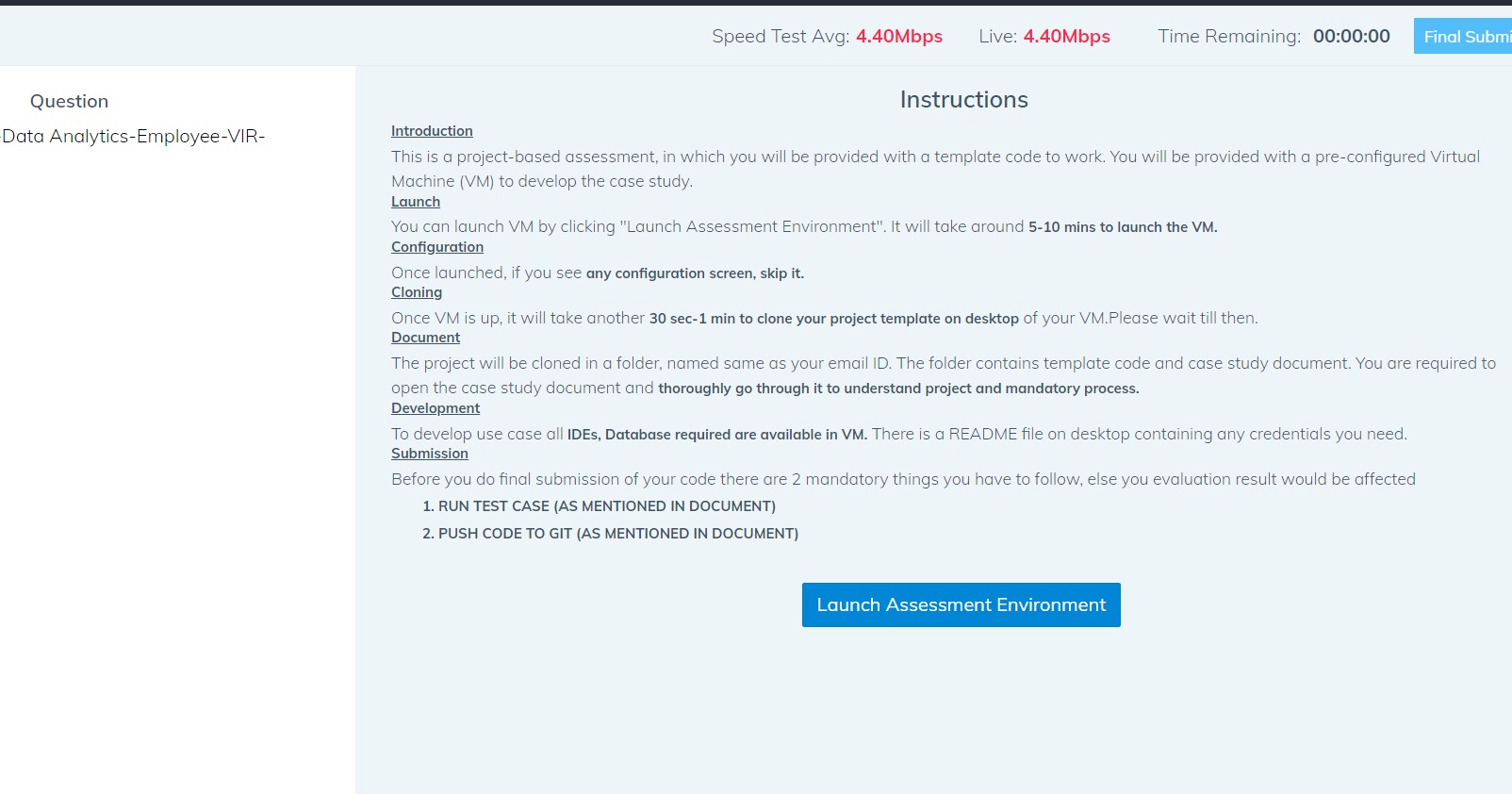
[fullstack@iiht.com](mailto:fullstack@iiht.com)

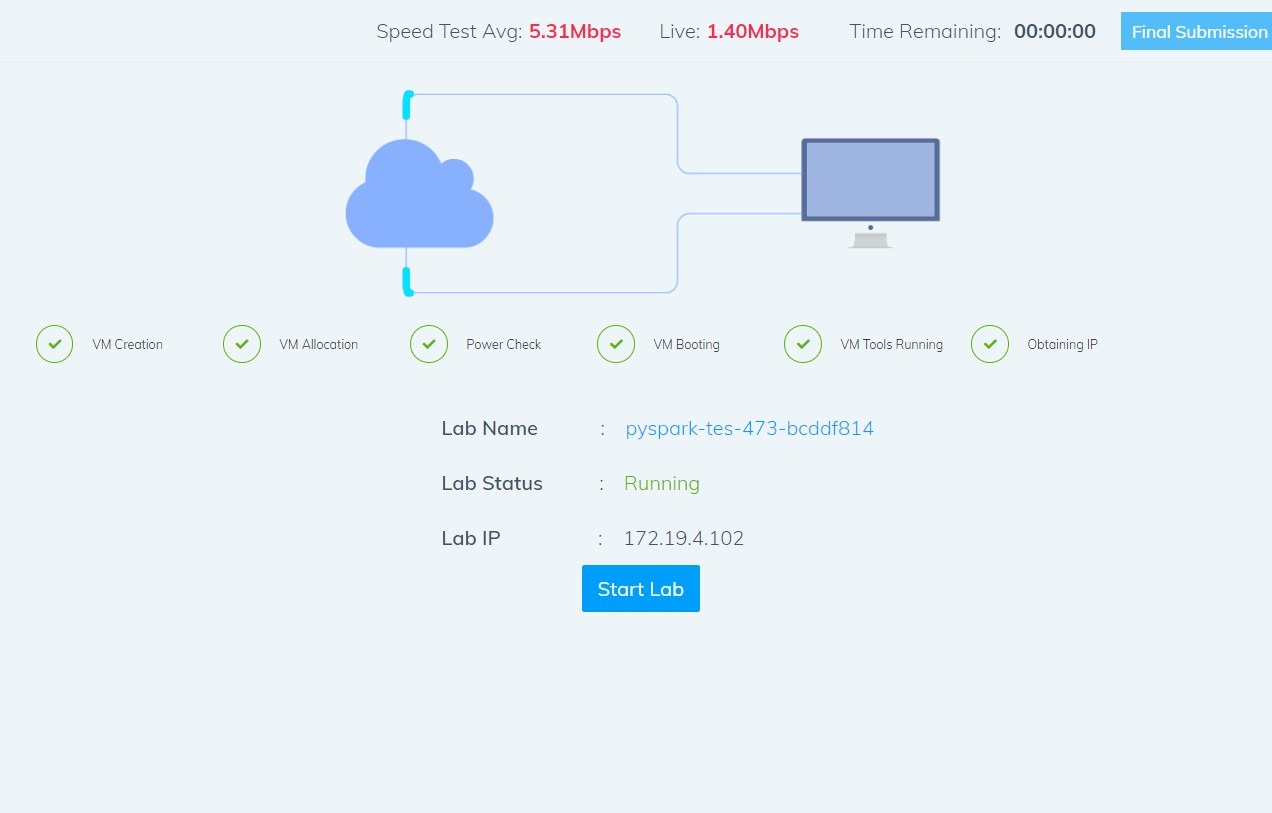
# Step to access the work environment

Step 1 use the URL to login provide the username and password



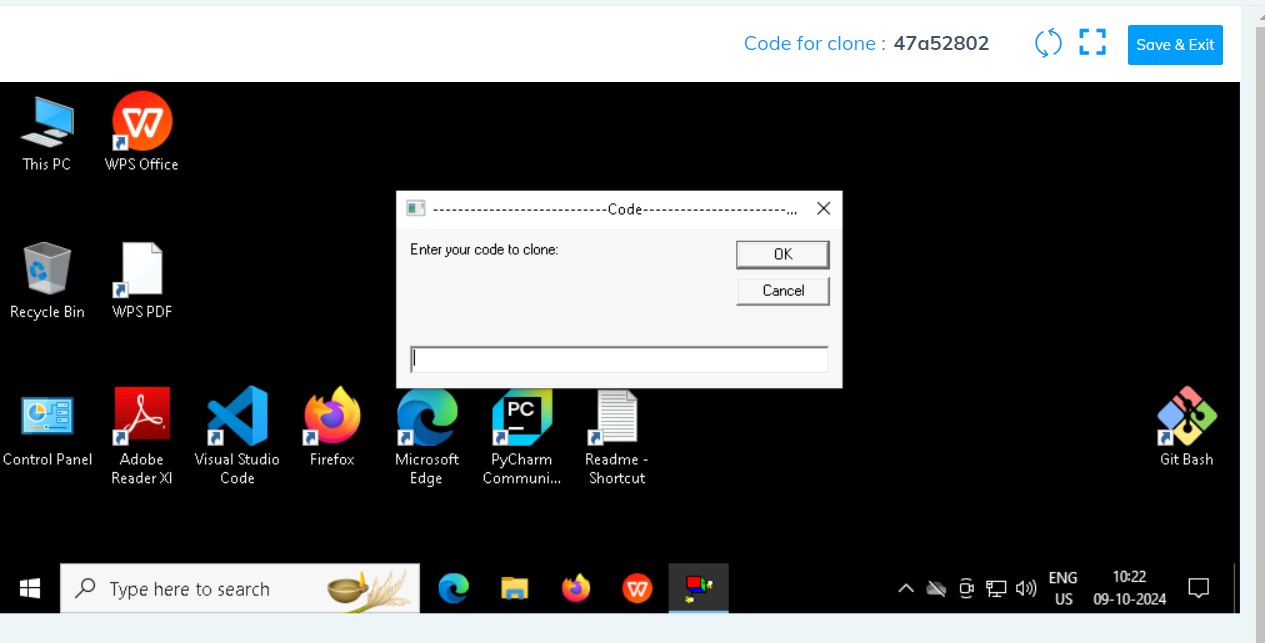
Step 2 Click on the launch assessment Environment





Step 3 Click on the start lab button

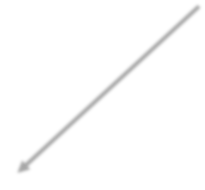
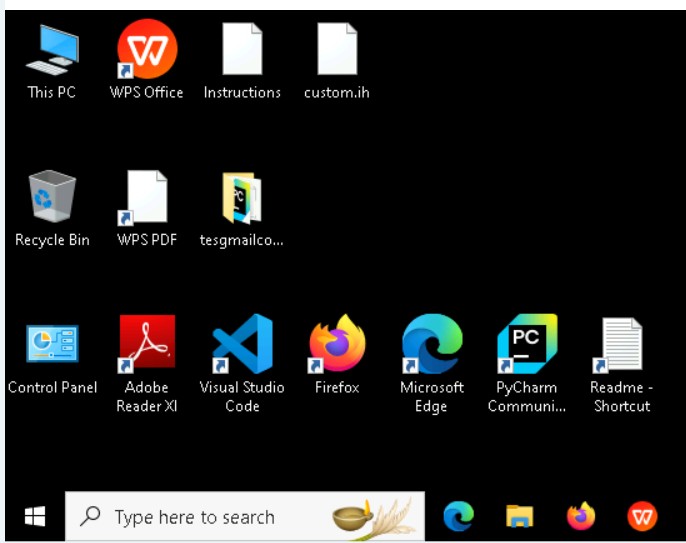
Step 4 you will get a window you need to type the code from that top corner



* You need to type the code in the window . It will take few minutes to start the window

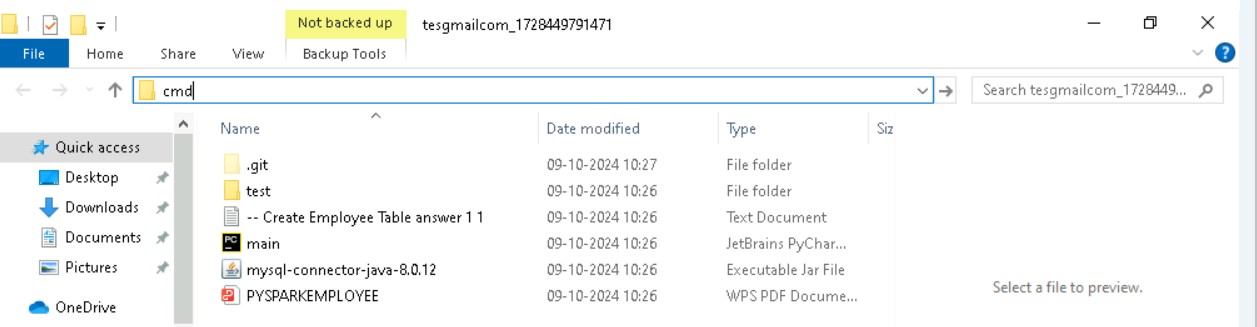
Click on ok

Step 5 after few seconds we can see that the your folder is cloned in the desktop .

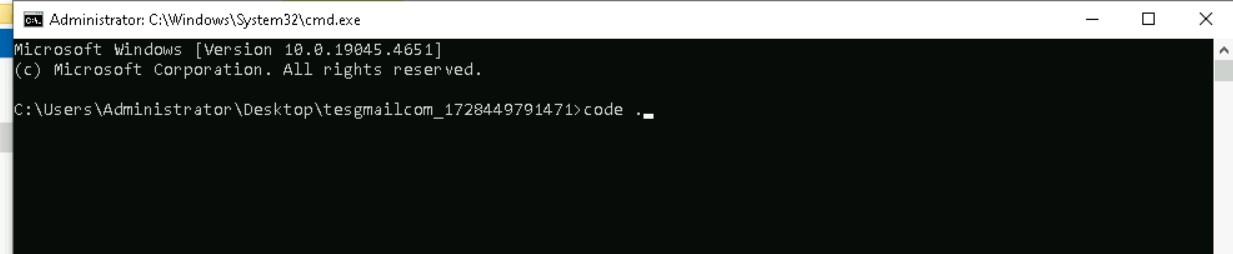


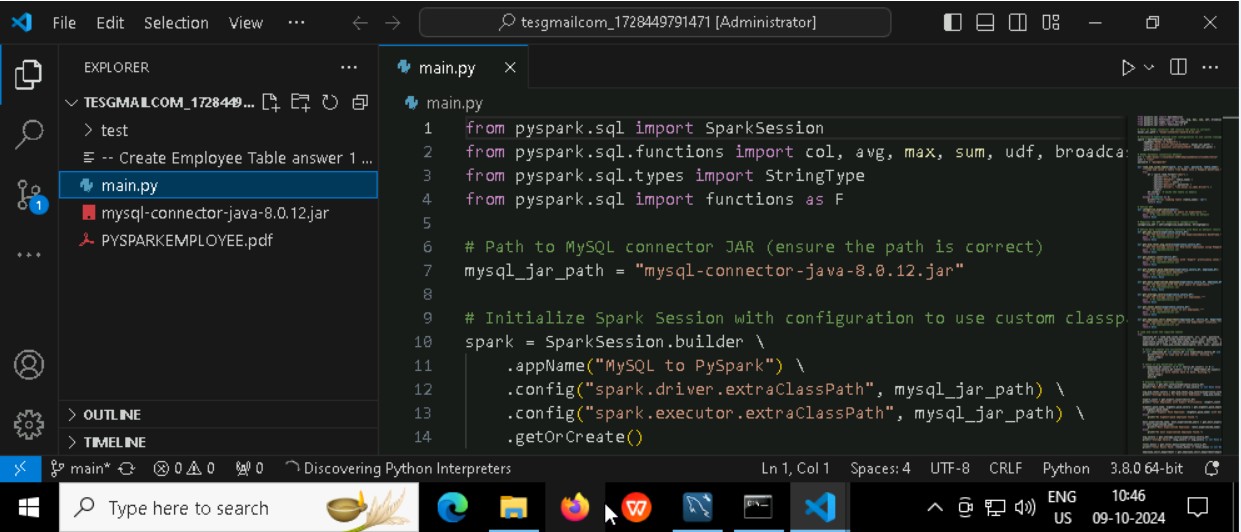
Folder cloned

Step 6 go inside the folder type cmd in the top of the file explorer



* Type code. And hit enter you can see that workspace is opened in the visual code





* You can see that workspace is ready to code

Note Please only work with visual code not with any other IDE

* In the folder cloned you will have all the project files needed .

Problem Statement : **Airline usecase for twitter**

Description : Use relevant methods operations toperform specified activities which are given in the instructions.

Abstract

The rise of social media platforms such as Twitter has revolutionized the way companies engage with their customers. Airlines, in particular, often receive real-time feedback on their services through tweets, providing valuable insights into customer sentiments. These insights can help airlines address customer concerns, enhance customer satisfaction, and improve their services. In this use case, we focus on performing sentiment analysis on airline-related tweets to classify them as positive, neutral, or negative, and to derive key insights from these sentiments.

Airlines are constantly monitoring social media for customer feedback, especially on platforms like Twitter. The ability to quickly identify the tone of these tweets allows airlines to respond to customer complaints and capitalize on positive sentiments. Sentiment analysis, using Natural Language Processing (NLP) techniques, is an effective way to categorize the emotional tone of a piece of text. By automating this process, airlines can streamline their customer service and marketing strategies.

This project uses a dataset of tweets related to different airlines and applies sentiment analysis techniques to classify these tweets. The analysis provides key performance indicators such as the most active airlines, the percentage of negative tweets, the airlines with the highest positive sentiment, and the most frequent phrases used by customers. By employing machine learning models, bigrams, and trigrams, we are able to analyze large volumes of social media data effectively.

**Techniques to be used to the usecase**

The primary technique for text preprocessing involves cleaning the tweet data by removing URLs, mentions, hashtags, and punctuation. This is followed by the application of the **TF-IDF Vectorizer** for converting text data into numerical vectors, which represent the importance of words in the dataset. The **Logistic Regression** model is then trained on these vectors to classify tweet sentiments as positive, negative, or neutral. Data splitting into training and testing sets is done using the **train\_test\_split** function from **sklearn**, ensuring model evaluation on unseen data.

For sentiment analysis, the code also includes additional exploratory data analysis (EDA) techniques. These involve calculating the sentiment distribution, identifying the most active airline, finding frequent n-grams like bigrams and trigrams, and analyzing tweet length. The **CountVectorizer** is used for n-gram extraction to uncover word pairs (bigrams) or triples (trigrams) that frequently appear within specific sentiment categories. Advanced metrics like sentiment ratios per airline are calculated by dividing sentiment counts by total counts per airline, providing insights into how specific airlines are perceived in terms of sentiment.

**Solve these Questions**

1. What is the sentiment distribution in the dataset?
2. Which airline is the most active in the dataset?
3. What percentage of tweets are negative in sentiment?
4. Which airlines have the highest positive sentiment?
5. What are the most frequent bigrams in positive sentiment tweets?
6. What is the average length of tweets in the dataset?
7. Which airline has the most negative sentiment associated with it?
8. Which airline has the highest positive sentiment ratio?
9. What are the most common trigrams in the dataset?
10. Which airlines have the highest neutral sentiment?
11. What are the most frequent words in negative sentiment tweets?

# Execution Steps to Follow:

1. All actions like build, compile, running application,running test cases will bethrough Command Terminal.
2. To open the command terminal the test takers, need to go to Application menu(Three horizontal lines at left top) -> Terminal -> New Terminal
3. This editor Auto Saves the code
4. If you want to exit(logout) and continue the coding later anytime (using Save & Exitoption on Assessment Landing Page) then you need to use CTRL+Shift+B-command compulsorily on code IDE. This will push or save the updated contents in the

internal git/repository. Else the code will not be available in the next login.

1. These are time bound assessments the timer would stop if you logout and while logging in back using the same credentials the timer would resume from the sametime it was stopped from the previous logout.
2. To setup environment:

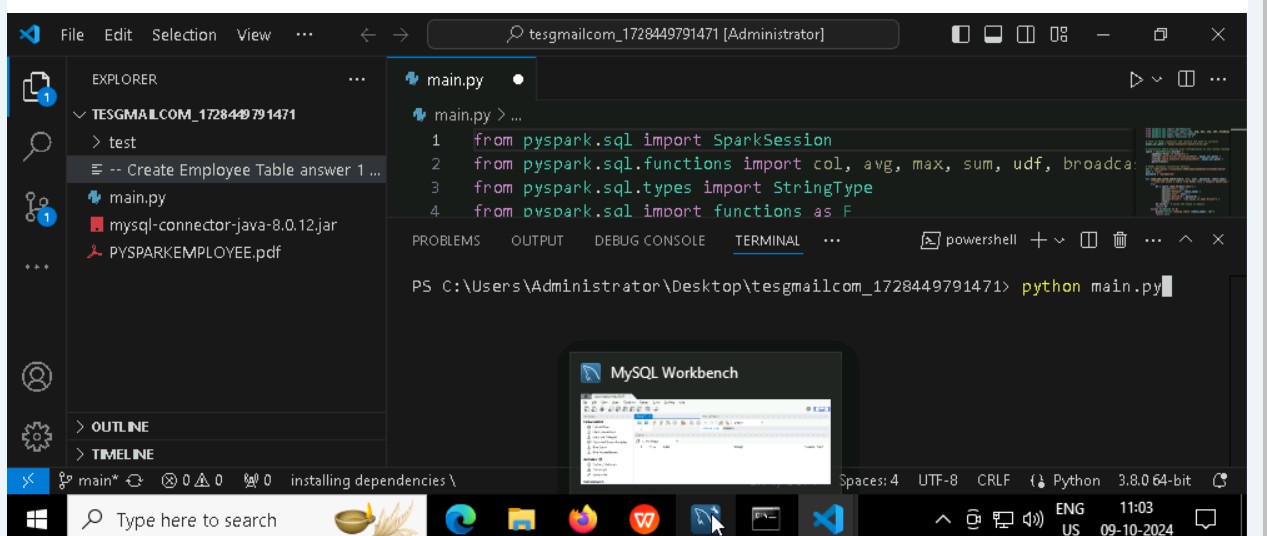
You can run the application without importing any packages

1. To launch application: Python usairline.py
2. To run Test cases:

python -m unittest

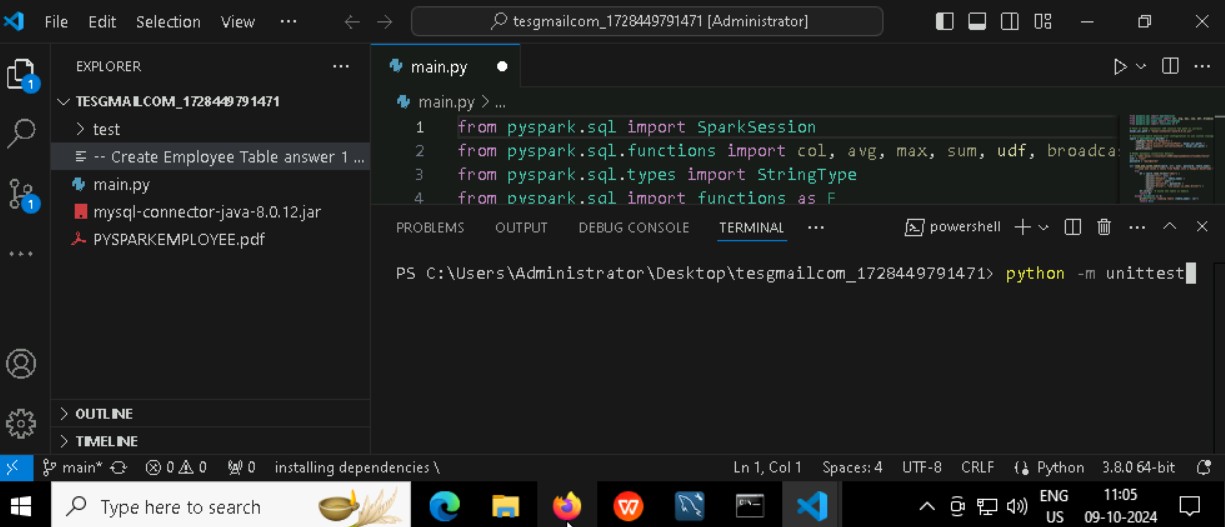
Before Final Submission also, you need to use CTRL+Shift+B- command compulsorily on code IDE. This will push or save the updated contents in the internalgit/repository for code

# Screen shot to run the program



**To run the application**

# Python football.py



**To run the testcase**

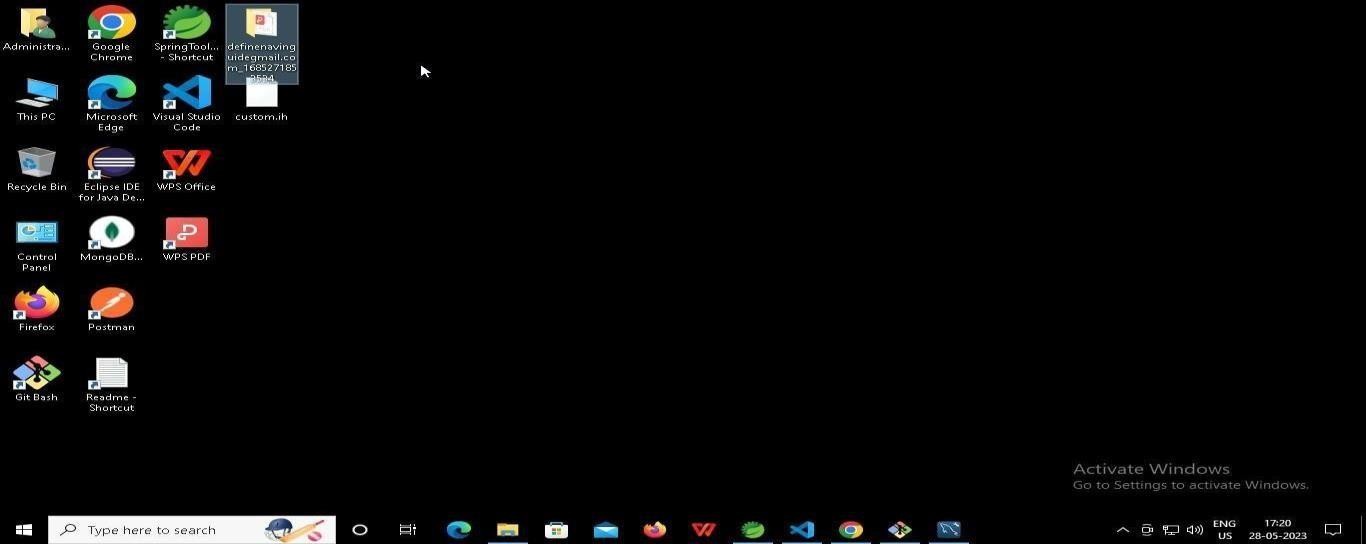
# Python -m unittest

**Screenshot to push the application to github**

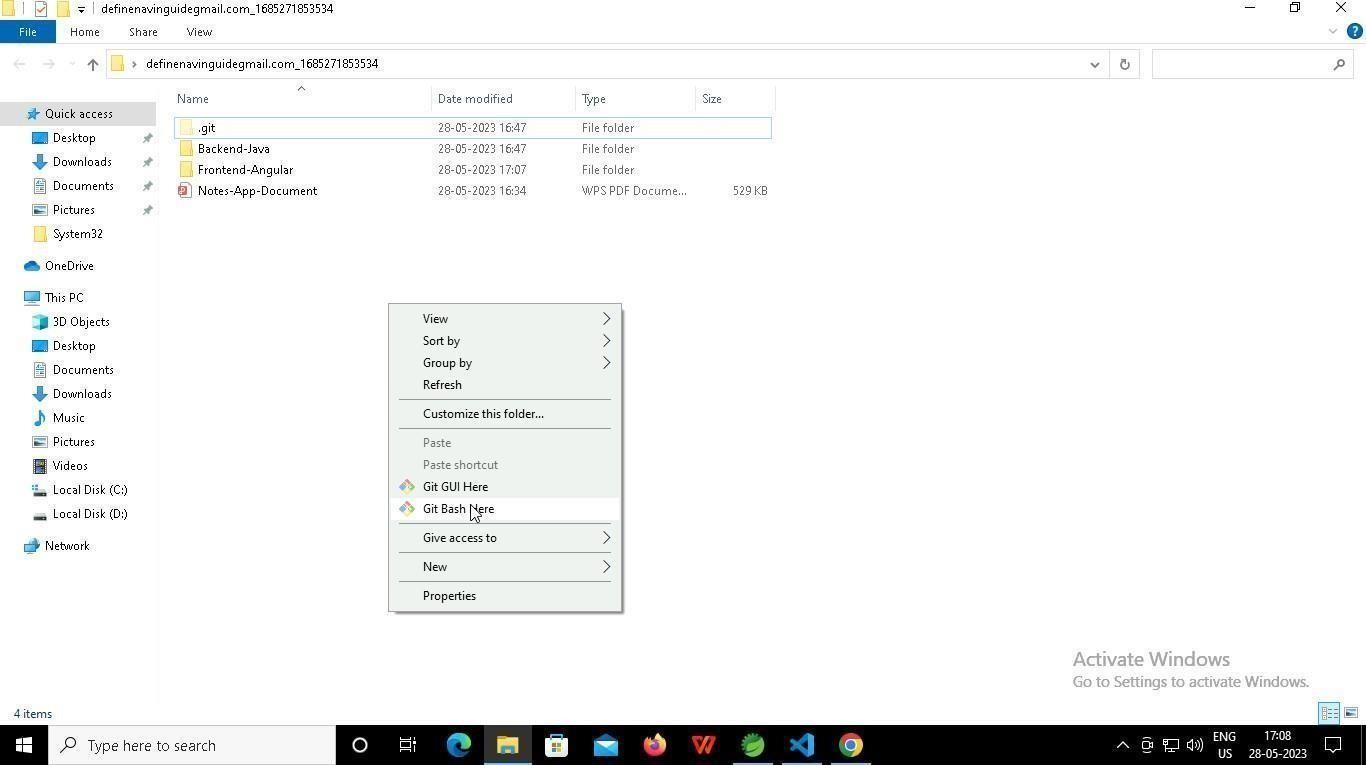
# -----x-----

**You can run test cases as many numbers of times and at any stage of Development, to check howmany test cases are passed/failed and accordingly refactor your code.**

1. **Make sure before final submission you commit all changes to git**. For that open theproject folder available on desktop

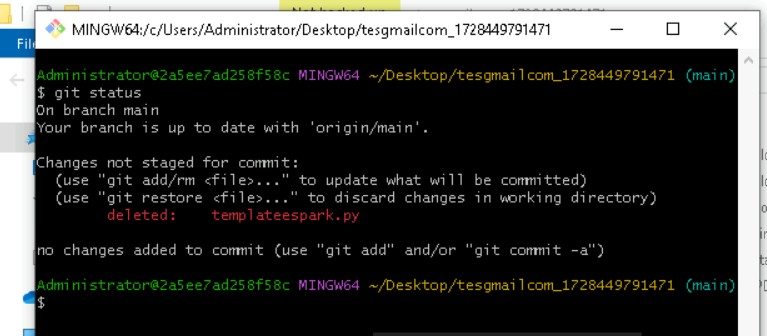


* 1. **Right click in folder and open Git Bash**



* 1. **In Git bash terminal, run following commands**

# git status

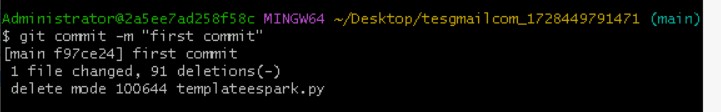


* 1. **git add .**

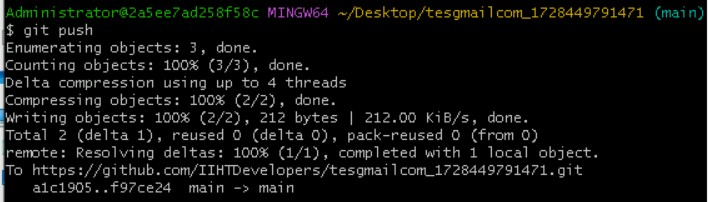


* 1. git commit -m “First commit”

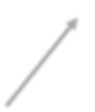
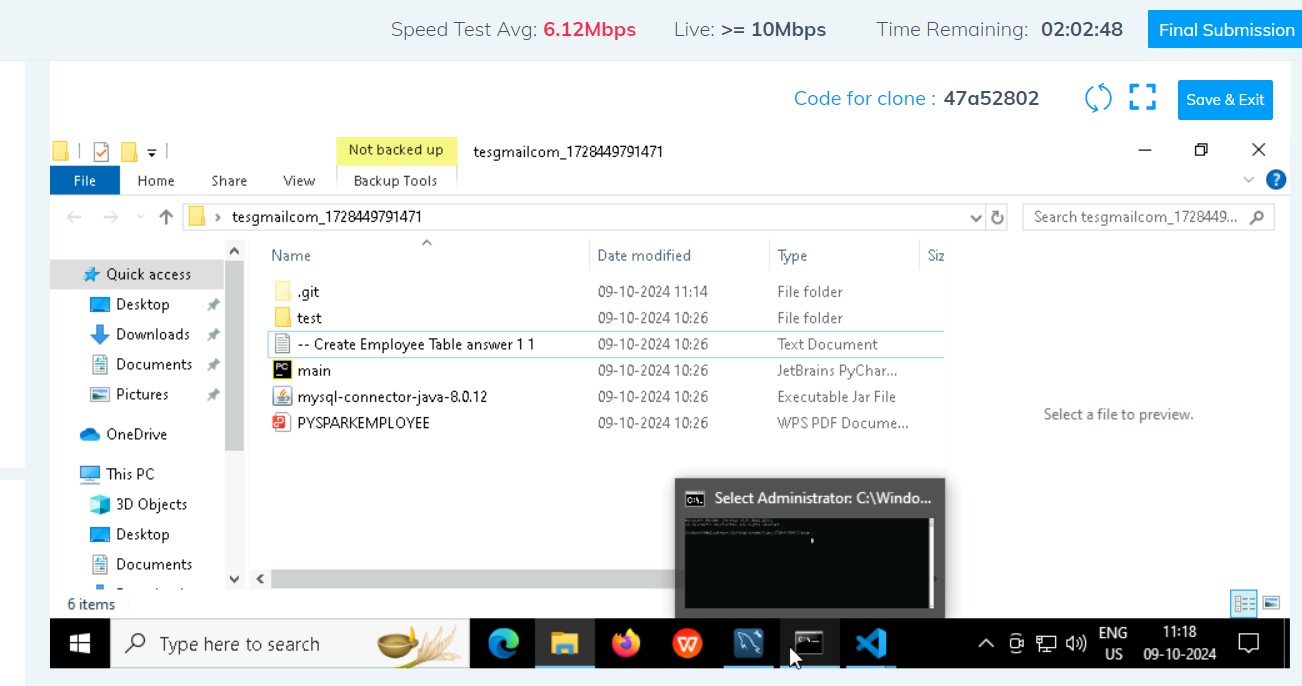
(You can provide any message every time you commit)



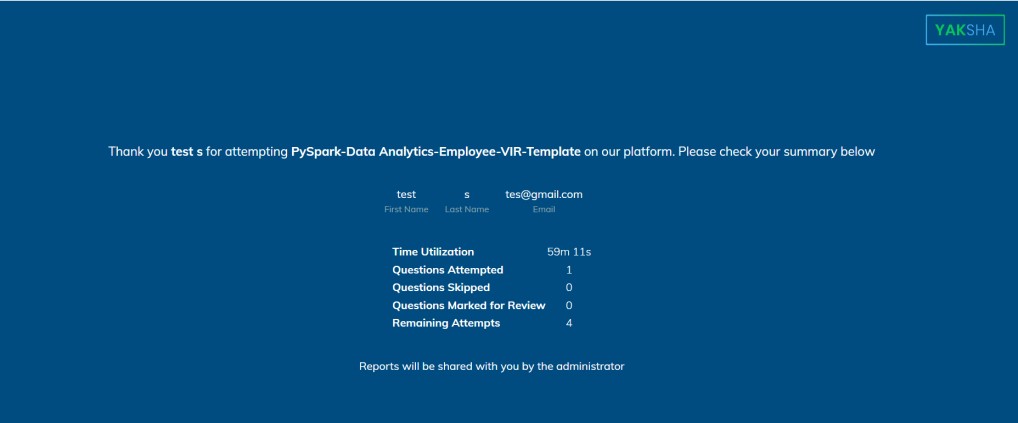
* 1. **git push**



# After you have pushed your code Finally click on the final submission button



**You should see a screen like this you will have to wait for the results . after getting this page you can leave the system**



# -----x-----