
System Requirements Specification Index

For

Kafka with python

Order processing system and analysis

Version 1.0

Problem Statement : Real-Time order processing and analysis

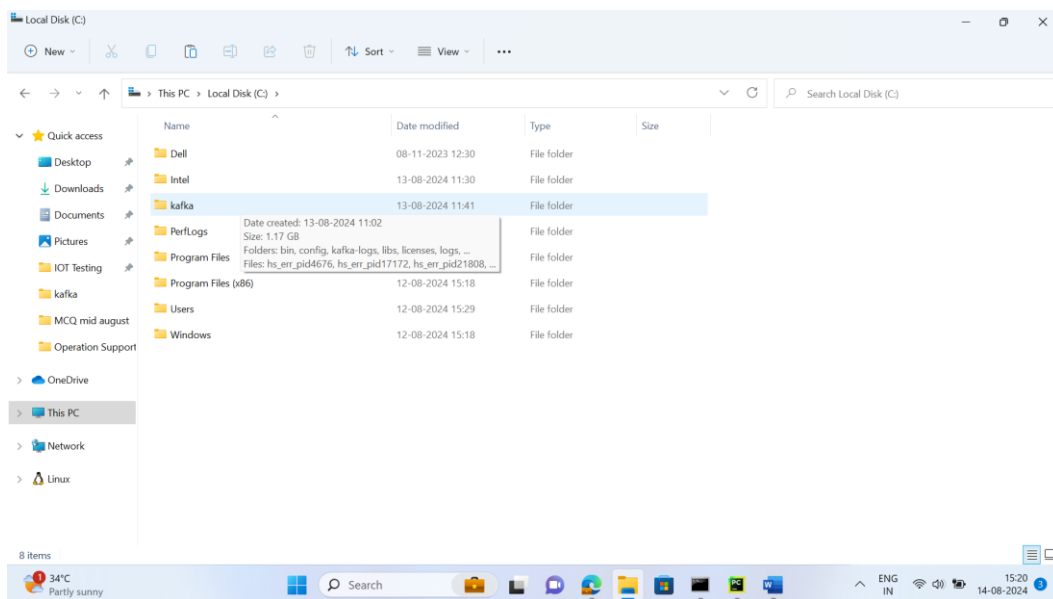
Description : Use relevant methods operations to perform specified activities which are given in the instructions.

Use case problem statement

An e-commerce application processes orders placed by customers. The orders contain product details like name, quantity, price, and category. The application is responsible for producing these orders as messages (perhaps for further processing in a message broker or downstream service) and for analyzing these orders to derive insights.

Kafka setup

You can see the kafka setup in the download folder



1. Extract Kafka:

- Extract the downloaded Kafka binaries into a directory of your choice (e.g., C:\kafka).

2. Set Up Zookeeper:

- Kafka uses Zookeeper to manage distributed brokers. Zookeeper is included in the Kafka distribution.

In the command prompt type the following command .

```
.\bin\windows\zookeeper-server-start.bat .\config\zookeeper.properties
```

```
C:\Windows\System32\cmd.exe - \bin\windows\zookeeper-server-start.bat .\config\zookeeper.properties
Microsoft Windows [Version 10.0.22000.2538]
(c) Microsoft Corporation. All rights reserved.

C:\kafka>.bin\windows\zookeeper-server-start.bat .\config\zookeeper.properties
[2024-08-13 11:37:25,658] INFO Reading configuration from: .\config\zookeeper.properties (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2024-08-13 11:37:26,191] INFO clientPortAddress is 0.0.0.0:2181 (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2024-08-13 11:37:26,191] INFO secureClientPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2024-08-13 11:37:26,191] INFO observerMasterPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2024-08-13 11:37:26,191] INFO metricsProvider.className is org.apache.zookeeper.metrics.impl.DefaultMetricsProvider (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2024-08-13 11:37:26,207] INFO autopurge.snapRetainCount set to 3 (org.apache.zookeeper.server.DataDirCleanupManager)
[2024-08-13 11:37:26,207] INFO autopurge.purgeInterval set to 0 (org.apache.zookeeper.server.DataDirCleanupManager)
[2024-08-13 11:37:26,207] INFO Purge task is not scheduled. (org.apache.zookeeper.server.DataDirCleanupManager)
[2024-08-13 11:37:26,207] WARN Either no config or no quorum defined in config, running in standalone mode (org.apache.zookeeper.server.quorum.QuorumPeerMain)
[2024-08-13 11:37:26,223] INFO Log4j 1.2 jmx support not found; jmx disabled. (org.apache.zookeeper.jmx.ManagedUtil)
[2024-08-13 11:37:26,223] INFO Reading configuration from: .\config\zookeeper.properties (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2024-08-13 11:37:26,238] INFO clientPortAddress is 0.0.0.0:2181 (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2024-08-13 11:37:26,238] INFO secureClientPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2024-08-13 11:37:26,238] INFO observerMasterPort is not set (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2024-08-13 11:37:26,238] INFO metricsProvider.className is org.apache.zookeeper.metrics.impl.DefaultMetricsProvider (org.apache.zookeeper.server.quorum.QuorumPeerConfig)
[2024-08-13 11:37:26,238] INFO Starting server (org.apache.zookeeper.server.ZooKeeperServerMain)
[2024-08-13 11:37:26,307] INFO ServerMetrics initialized with provider org.apache.zookeeper.metrics.impl.DefaultMetricsProvider@22e357dc (org.apache.zookeeper.server.ServerMetrics)
[2024-08-13 11:37:26,333] INFO ACL digest algorithm is: SHA1 (org.apache.zookeeper.server.auth.DigestAuthenticationProvider)
[2024-08-13 11:37:26,333] INFO zookeeper.DigestAuthenticationProvider.enabled = true (org.apache.zookeeper.server.auth.DigestAuthenticationProvider)
```

In an another command prompt

Start Kafka Broker:

.\bin\windows\kafka-server-start.bat .\config\server.properties

```
C:\Windows\System32\cmd.exe - \bin\windows\kafka-server-start.bat .\config\server.properties
Microsoft Windows [Version 10.0.22000.2538]
(c) Microsoft Corporation. All rights reserved.

C:\kafka>.bin\windows\kafka-server-start.bat .\config\server.properties
[2024-08-13 11:41:07,496] INFO Registered kafka:type=kafka.Log4jController MBean (kafka.utils.Log4jControllerRegistration$)
[2024-08-13 11:41:09,315] INFO Setting -D jdk.tls.rejectClientInitiatedRenegotiation=true to disable client-initiated TLS renegotiation (org.apache.zookeeper.common.X509Util)
[2024-08-13 11:41:09,407] INFO RemoteLogManagerConfig values:
    log.local.retention.bytes = -2
    log.local.retention.ms = -2
    remote.fetch.max.wait.ms = 500
    remote.log.index.file.cache.total.size.bytes = 1073741824
    remote.log.manager.copier.thread.pool.size = 10
    remote.log.manager.copy.max.bytes.per.second = 9223372036854775807
    remote.log.manager.copy.quota.window.num = 11
    remote.log.manager.copy.quota.window.size.seconds = 1
    remote.log.manager.expiration.thread.pool.size = 10
    remote.log.manager.fetch.max.bytes.per.second = 9223372036854775807
    remote.log.manager.fetch.quota.window.num = 11
    remote.log.manager.fetch.quota.window.size.seconds = 1
    remote.log.manager.task.interval.ms = 30000
    remote.log.manager.task.retry.backoff.max.ms = 30000
```

Create partitions with the count of 2 #[important step]#

To create partitions

kafka-topics.bat --create --topic order_topic--bootstrap-server localhost:9092 --partitions 2 --replication-factor 1

```
C:\kafka\bin\windows>kafka-topics --describe --topic order_topic --bootstrap-server localhost:9092
[2024-09-19 18:32:43,445] WARN [AdminClient clientId=adminclient-1] The DescribeTopicPartitions API is not supported, using Metadata API to describe topics. (org.apache.kafka.clients.admin.KafkaAdminClient)
Topic: order_topic      TopicId: vk7urQn1TXamH65Vh6NC-w PartitionCount: 2      ReplicationFactor: 1      Configs:
  Topic: order_topic      Partition: 0      Leader: 0      Replicas: 0      Isr: 0 Elr: N/A      LastKnownElr: N/A
  Topic: order_topic      Partition: 1      Leader: 0      Replicas: 0      Isr: 0 Elr: N/A      LastKnownElr: N/A
```

Question using the template provide your are requested to create python code

Implement the logic Use[order_id]as the key to control which partition the message goes toSend even order_ids to one partition, odd to another

- Calculate how many messages are sent to the consumer
- What is the Total revenue from orders ?
- Which product has the highest revenue for the business
- What is the average price of electronic items?
- How many personal care products are available in the dataset .
- What is the standard deviation of prices.
- Identify the produce with the minimum-price.

Execution Steps to Follow:

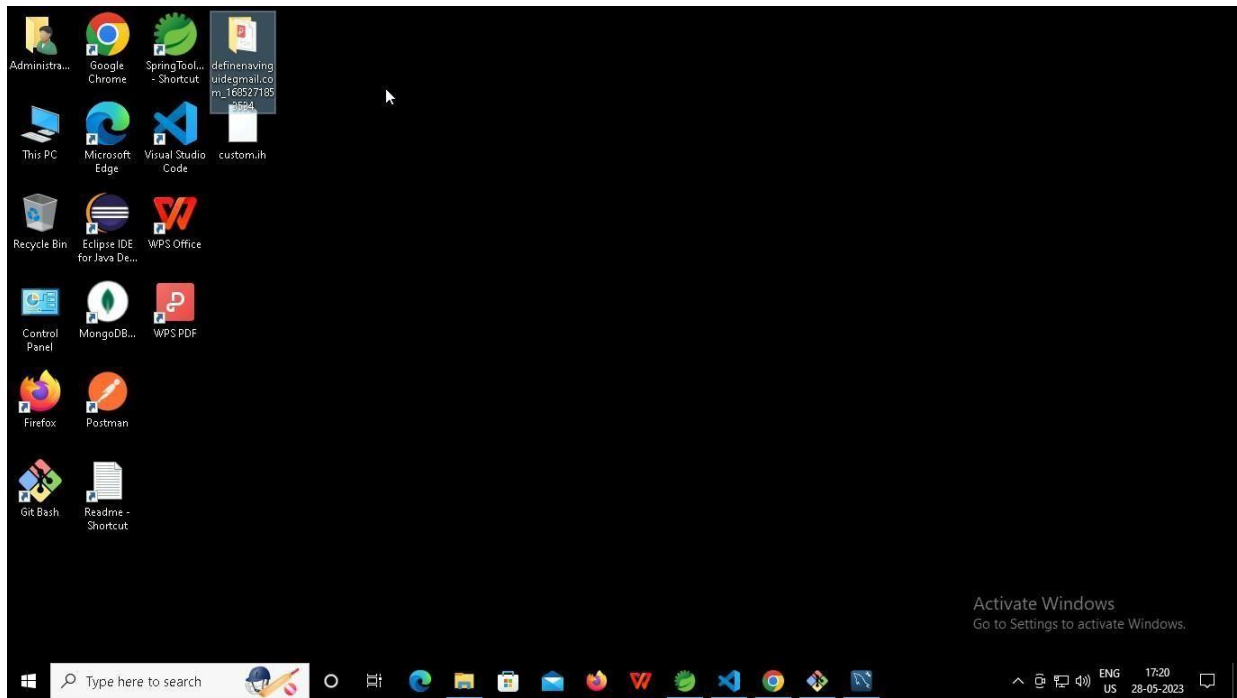
1. Use the same dataset in Json file
2. Capture the data from producer to consumer
3. Perform all the methods
4. Upload the code to the Github
5. Submit the code the get the test auto graded

Python libraries to be installed

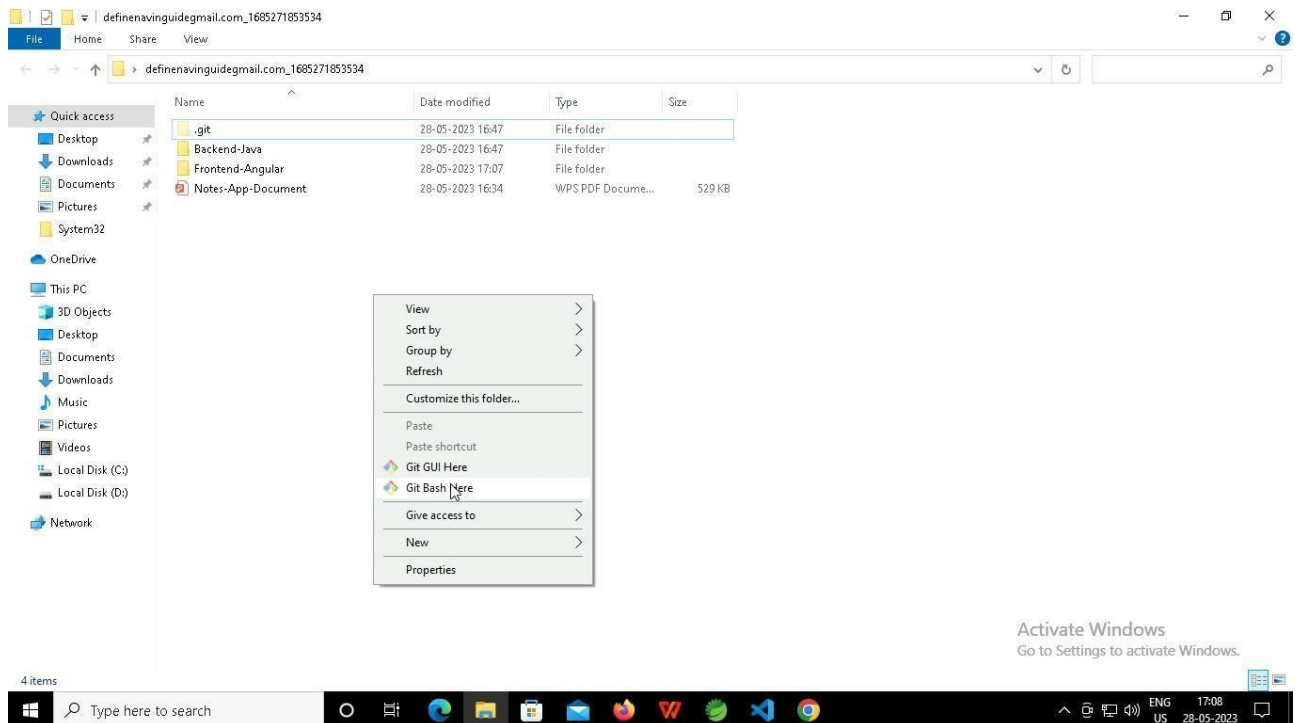
- pip install confluent_kafka
- pip install requests
- To run the unittest for the program
- Python -m unittest

Steps to upload the code to github

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

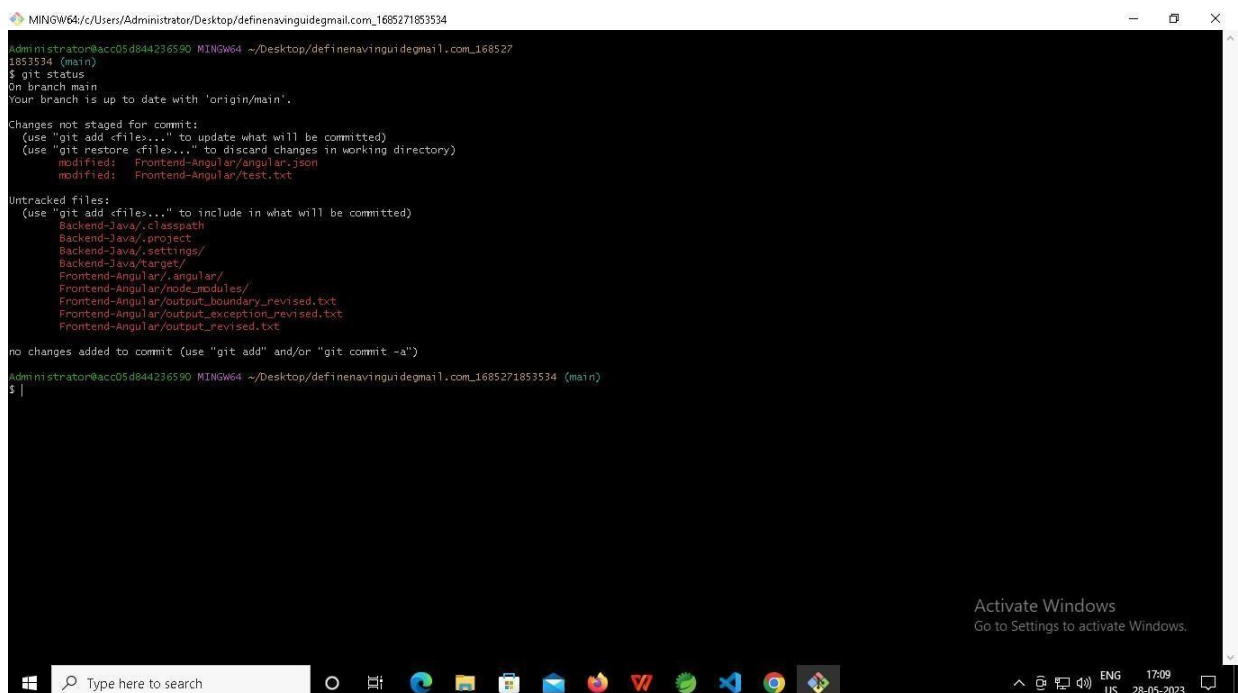


a. Right click in folder and open Git Bash

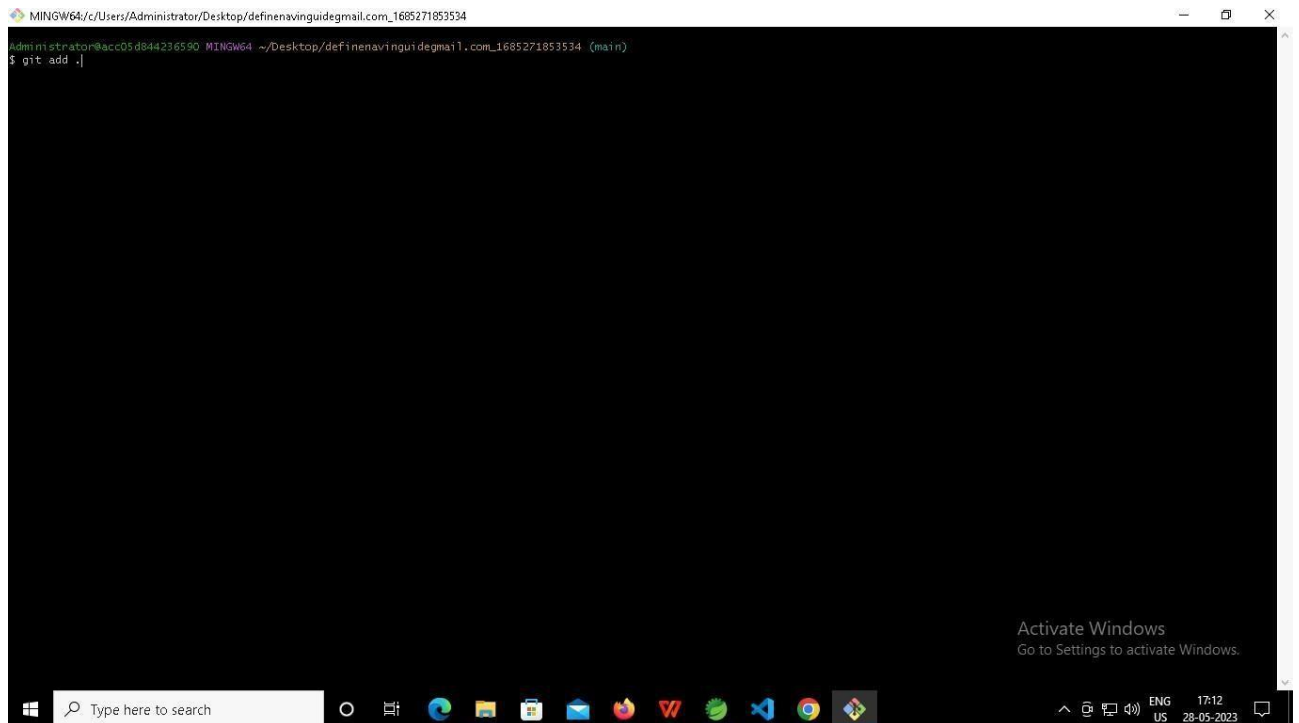


b. In Git bash terminal, run following commands

c. git status

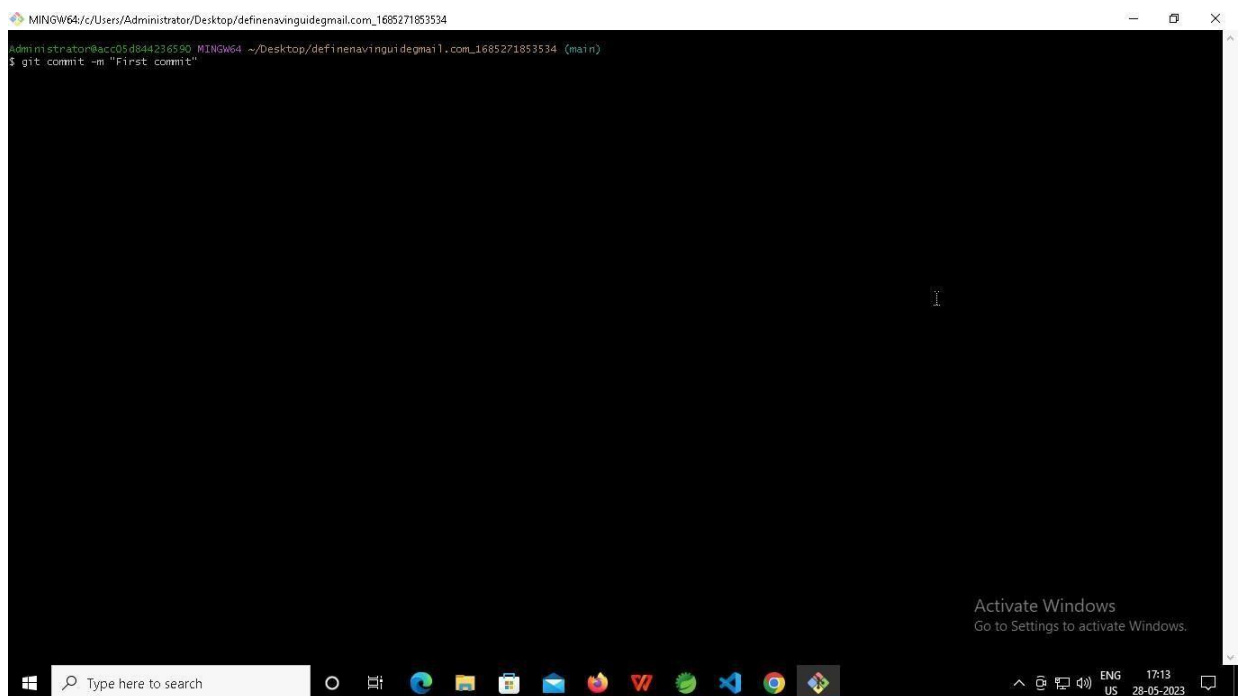


d. `git add .`



A screenshot of a Windows terminal window titled "MINGW64/c/Users/Administrator/Desktop/definavinguidegmail.com_1685271853534". The terminal shows the command prompt with the user "Administrator@ecc05d844236590" and the directory "~/Desktop/definavinguidegmail.com_1685271853534 (main)". The command `$ git add .` has been entered. The terminal background is black with white text. At the bottom, there is a Windows taskbar with various application icons and a system tray showing the date and time as 17:12 on 28-05-2023. An "Activate Windows" watermark is visible in the bottom right corner of the terminal window.

e. `git commit -m "First commit"`
(You can provide any message every time you commit)



A screenshot of a Windows terminal window titled "MINGW64/c/Users/Administrator/Desktop/definavinguidegmail.com_1685271853534". The terminal shows the command prompt with the user "Administrator@ecc05d844236590" and the directory "~/Desktop/definavinguidegmail.com_1685271853534 (main)". The command `$ git commit -m "First commit"` has been entered. The terminal background is black with white text. At the bottom, there is a Windows taskbar with various application icons and a system tray showing the date and time as 17:13 on 28-05-2023. An "Activate Windows" watermark is visible in the bottom right corner of the terminal window.

f. git push

```
MINGW64/c/Users/Administrator/Desktop/definavinguidegmail.com_1685271853534
Administrator@acc05d844236590 MINGW64 ~/Desktop/definavinguidegmail.com_1685271853534 (main)
$ git push
Enumerating objects: 28156, done.
Counting objects: 100% (28156/28156), done.
Delta compression using up to 4 threads
Compressing objects: 100% (19230/19230), done.
Writing objects: 100% (28151/28151), 71.20 MiB | 3.27 MiB/s, done.
Total 28151 (delta 7783), reused 28148 (delta 7780), pack-reused 0
remote: Resolving deltas: 100% (7783/7783), completed with 3 local objects.
remote: warning: See https://git.io/lfs for more information.
remote: warning: File Frontend-Angular/.angular/cache/15.2.8/angular-webpack/9b29720b052dd6ef081cbdb91ce119f5ada5942/0.pack is 86.24 MB; this is larger than GitHub's recommended maximum file
size of 50.00 MB
remote: warning: GH001: Large files detected. You may want to try Git Large File Storage - https://git-lfs.github.com.
To https://github.com/Evaluate2/definavinguidegmail.com_1685271853534.git
  490cd2db..610226ed main -> main

Administrator@acc05d844236590 MINGW64 ~/Desktop/definavinguidegmail.com_1685271853534 (main)
$ |
```

Activate Windows
Go to Settings to activate Windows.

Type here to search

ENG US 17:15 28-05-2023

-----X-----