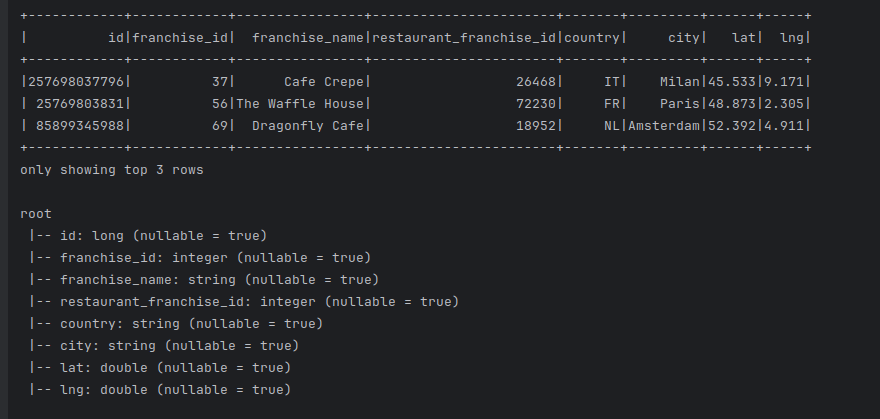
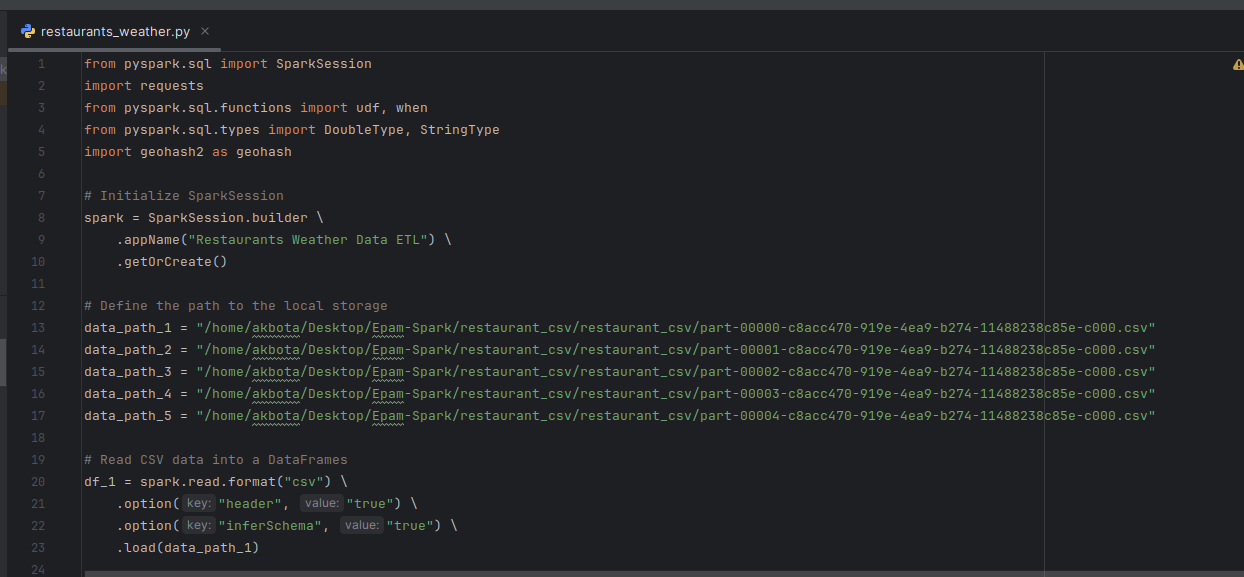
Screenshots. Comments. Code documentation.

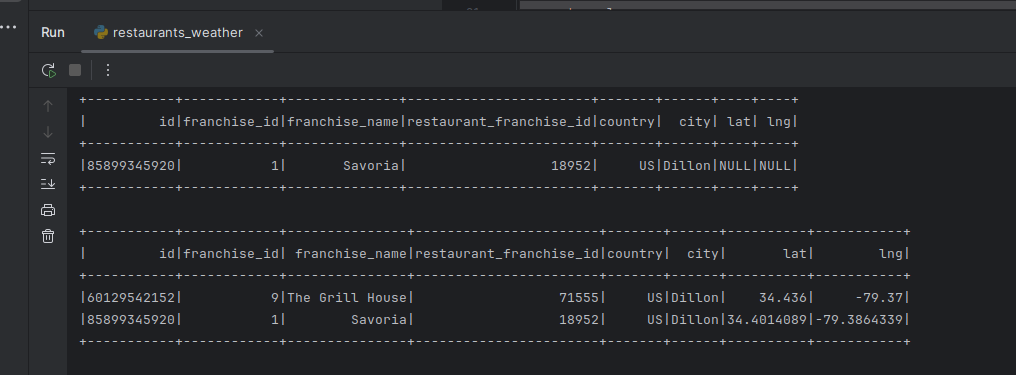
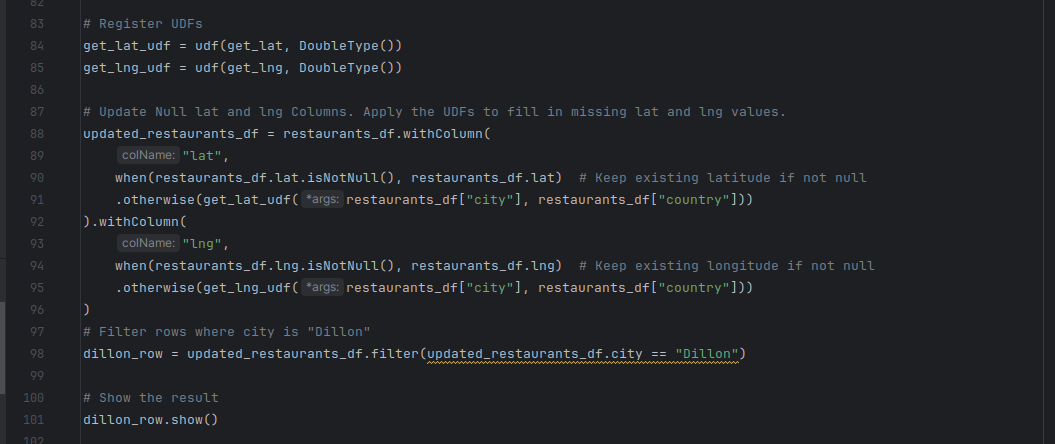
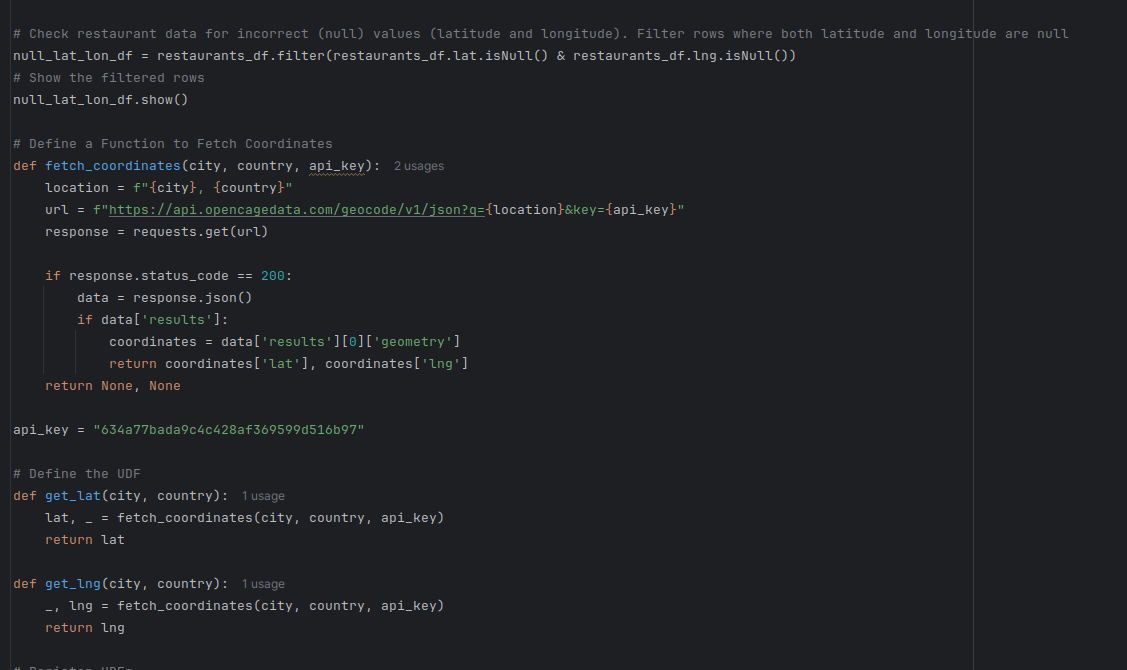
1. **Install Spark locally using one of the methods described [here](https://spark.apache.org/downloads.html) or in Docker.**

I installed PyCharm IDE on my computer. There is special package called **pyspark** in IDE and I installed it. Another installations were related to modules **requests** to work with remote system via REST API, **geohash** to calculate hash of geographical places. These modules were installed by **pip install** command in terminal.

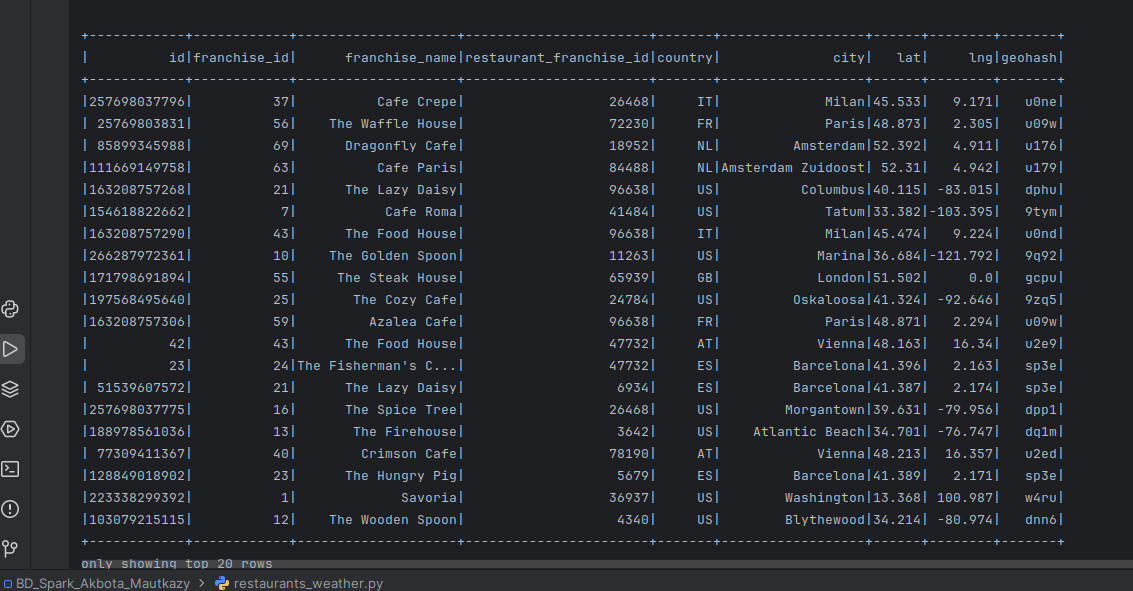
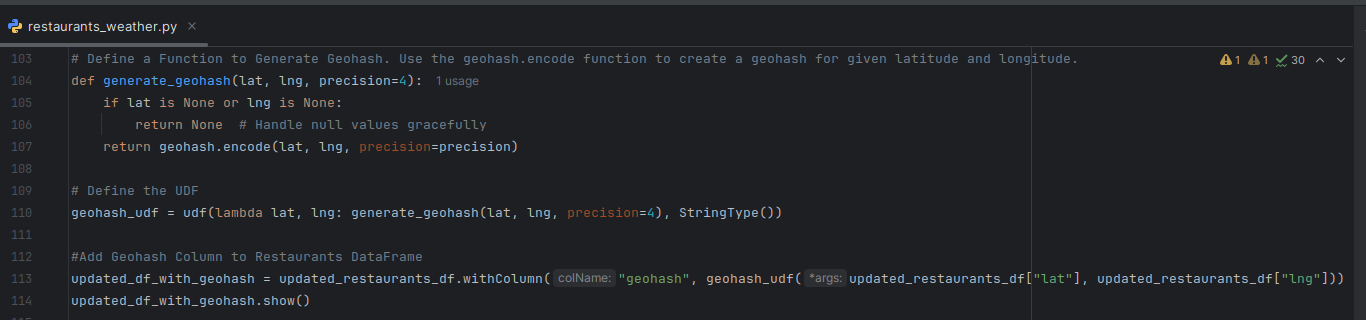
**2. Create a Spark ETL job to read data from a local storage.**



**3. Check restaurant data for incorrect (null) values (latitude and longitude). For incorrect values, map latitude and longitude from the OpenCage Geocoding API in a job via the REST API.**

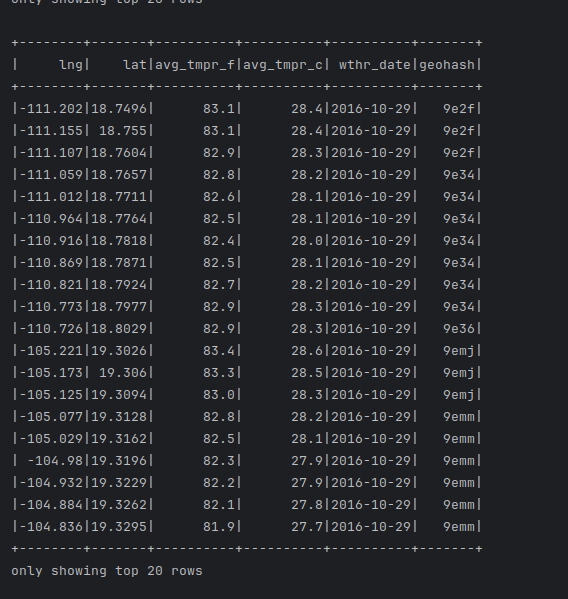
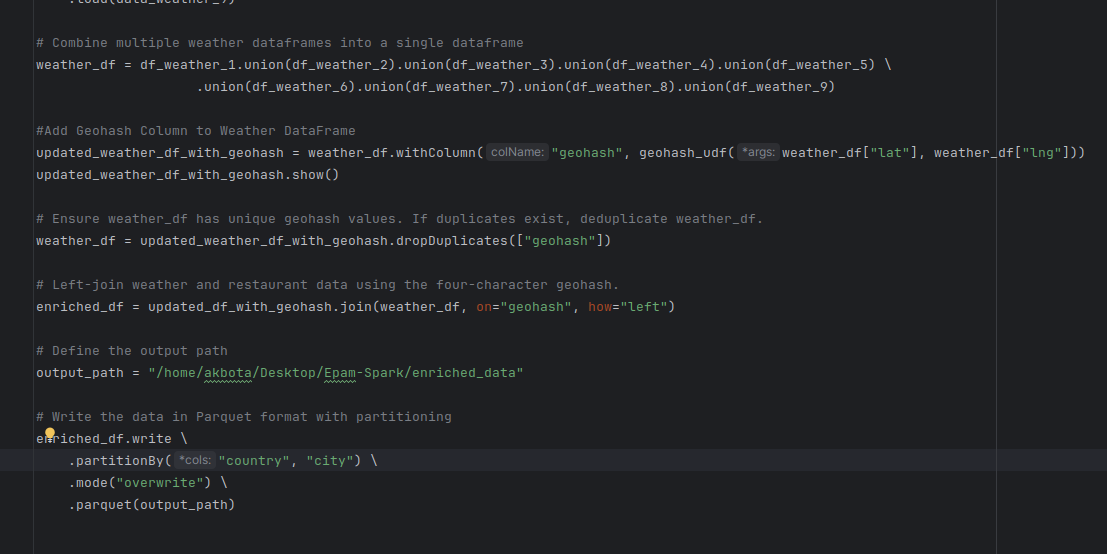
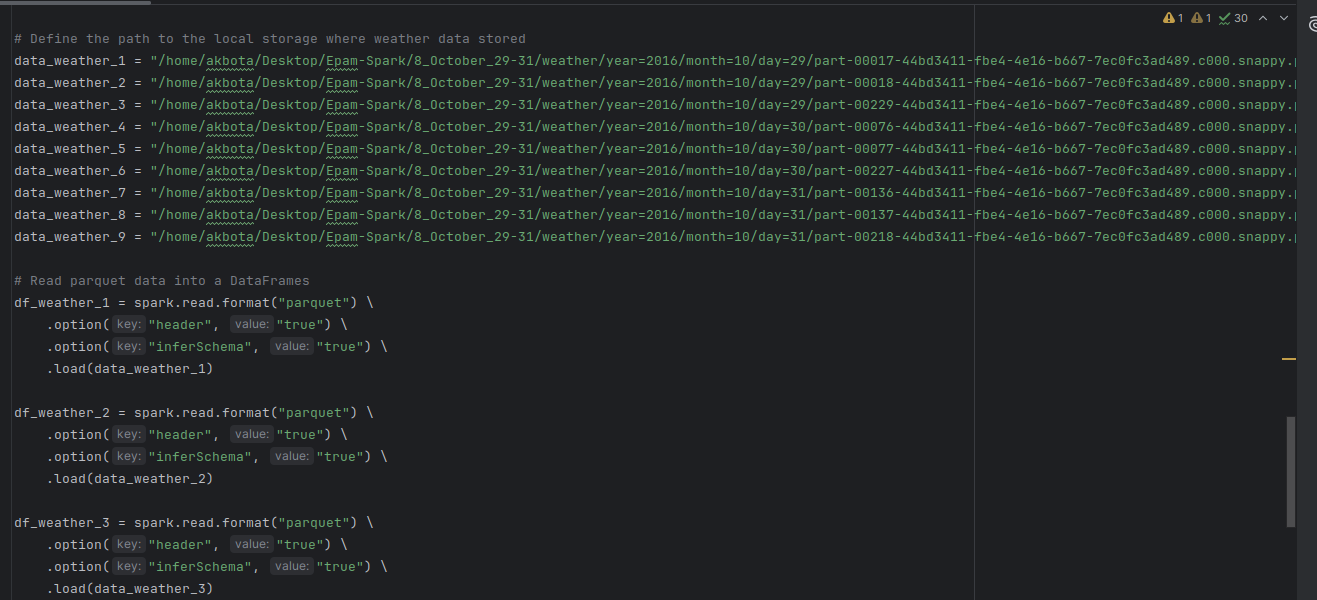


4. **Generate a geohash by latitude and longitude using a geohash library like geohash-java. Your geohash should be four characters long and placed in an extra column.**



5. **Left-join weather and restaurant data using the four-character geohash. Make sure to avoid data multiplication and keep your job idempotent.**

**Store the enriched data (i.e., the joined data with all the fields from both datasets) in the local file system, preserving data partitioning in the parquet format.**



**7. Unit test**

