Mini Project #1

Kaggle link: https://www.kaggle.com/code/cckulinda/mini-project-enterprise

Objective: Convert data semi-structured (Yelp data in JSON) to structured (SQL Server on Azure) with proper metadata specification.

Assessment - 20 points - Teamwork

Skills: Read and Interpret JSON Data/ Convert JSON to MySQL / Update MySQL tables by accessing API's

Instructions

1. Create an account on Kaggle(need not use your Illinois email, but you can. Use the Yelp Dataset on KaggleLinks to an external site. and 'Create new notebook'. Each team member should do this and get practice. If you want to work on one notebook, you can 'share' and 'Add collaborators'. When it is time for final submission, one team member can share the final notebook in the assignment. A notebook can have collaborators so you can all edit the final notebook if needed. We are using Kaggle so you don't need to download and upload the data again to Google, and it is also a great place to start participating in analytics competitions. If you get placed in some top 10/50/100 lists, that is very good for your interviews.

2. Explore the JSON data in the notebook, or convert it to a Pandas Dataframe and explore the data. Understand the structure of JSON data files. Pay particular attention to the time period for which you have data and the scope of the data(location/category etc). Document all code and metadata in the Kaggle Notebook. This is exploratory analysis, and is important. This notebook must demonstrate that you understand the data.

3. Convert the JSON data so you can write the data into a database you create on Azure SQL Server database. See steps/video at the end. All the tables should be created and data ported over. Metadata should be set appropriately. You can reference the yelp_champaign data model to get the correct metadata specifications. This should all be done in kaggle/Google Collab as needed. Kaggle and Google Collab have the same functionality. Since the data is on

Kaggle, I recommend staying with Kaggle.

4. Now for the Business Value from this exercise: You want to set up a new restaurant, and you have yelp data. Write 10 queries on the yelp data which help you find answers to questions that will help you start a new restaurant in a city of your choice. It must be clear from the queries

5. Submit all of this as public Kaggle notebooks/Google collab notebooks. One submission per team. Discuss resources on the forum. Help each other.

6. After you submit, review two other submissions anonymously and give suggestions/comments to the other team within 24 hours of the due date.

7. Fill out a peer review survey that helps you rate your own work and the work of your team mates. This impacts your grade, so make sure you contribute to the project and help your team mates contribute. This is a team effort.

Steps to create SQL server and database (Azure) - videoLinks to an external site.

Additional Tips - tips

Some Rubric

Some Rubric

Criteria	Ratings	Pts
This criterion is linked to a Learning OutcomeLoading and exploring the JSON data in the notebook		1 pts
This criterion is linked to a Learning OutcomeConverting JSON data to a Pandas DataFrame and further exploration		1 pts
This criterion is linked to a Learning OutcomeCode written to explore the structure, time period, and scope of the data		1 pts
This criterion is linked to a Learning OutcomeComprehensive		2 pts

Some Rubric

Criteria	Ratings	Pts
documentation and commenting on the code		
This criterion is linked to a Learning OutcomeConverting JSON data for MySQL database compatibility		1 pts
This criterion is linked to a Learning OutcomeCreating all tables in Azure		2 pts
This criterion is linked to a Learning OutcomeAll data in all tables ported over		2 pts
This criterion is linked to a Learning OutcomeBreak apart multi valued attributes into tables		2 pts
This criterion is linked to a Learning OutcomeSetting appropriate metadata, referencing yelp_champaign database specifications		2 pts
This criterion is linked to a Learning OutcomeFormulating 10 SQL queries relevant to starting a new restaurant (0.3 points for each query)		3 pts
This criterion is linked to a Learning OutcomeQueries providing insightful information for decision-making (0.3 points for each query)		3 pts
Total Points: 20		