The Database:

* Hosted instance on Amazon Web Services, access through secure link with PGAdmin through a PostgreSQL database
* PostgreSQL allows the entire team access to the database on their own terminal, reducing the need for data transfers

Interfacing:

* The webscraping file includes a connection string that imports the data directly to PostrgreSQL and, by proxy, to AWS
* The machine learning model also includes a connection string that imports data directly from PostgreSQL

Info Page:

* Authored the Info page – zodiac storytime!

Challenges & Successes:

* Planning – better planning, ie – data types at each step, format of data or documents when they are to be passed to another team member, etc
  + Prevents overlapping labor
  + Provides better structure and expectations
* Learning by doing – gaining a perspective of real project-driven scenarios
  + Better insight into what questions need to be asked
  + Better techniques to process information in the quest for understanding

Blurb:

Database:

Once we had found our data, we needed to organize it and store it. For storage we opted for hosting on AWS. The Amazon RDS (Relational Database Service) provides dynamic storage that interfaces directly with PostgreSQL. With a connection to the host, all team members could access the data in PGAdmin through a PostgreSQL database. Although most of the cleaning and organizing was done primarily with Pandas and Python but a small amount was also accomplished in PostgreSQL. It’s is not only helpful for the team members to have direct access to the data, it also removes the need to send large data files by email.

Once the database was built and connected to AWS, tables were built and data was loaded as a test. An inner join was used to combine data from separate tables. A connection string was created. Using a connection engine in the the webscraping file and the machine learning file, the database was integrated into the ETL.

Challenges

Planning was the most important thing we could have done differently. I caught myself doing work that didn’t need to be done or had been done. There were also some junctures in the work that could have been planned out better.

Ex 1. Cleaning data that got removed.

Ex 2. Cleaning data that is cleaned.

Ex. 3 Clearly defined formats for all work that is to be passed to another team member.

Successes

Learning to have some insights into the workflow. Knowing more in-depth questions to think about in the planning phase of analytics projects.

Info Page:

I authored the zodiac information. As I sifted through the signs and their traits, I decided to highlight the negative traits and make them all killers. Otherwise, why are we even talking about this!

Diagram

Description automatically generated

Graphical user interface, text, application, chat or text message

Description automatically generated