CS 586 Final Project, Part 1

Dylan Laufenberg

1 Data Domain

The data source I will use is the NewSpace Index (https://www.newspace.im/index.html), which describes space activities related to the NewSpace group. This website presents a "concise overview of new commercial satellite constellations, small satellite rocket launchers, most exciting space missions, future spaceships and space tourism opportunities."

There are nine tables on the website, with the following sizes (columns x rows):

1. Constellations — 7 x 90
2. Launchers — 11 x 88
3. Funding — 5 x 16
4. Upcoming Spaceships — 4 x 7
5. Small Reentry Vehicles — 5 x 7
6. NewSpace Missions — 5 x 6
7. Interesting Companies — 5 x 16
8. Private Space Stations — 5 x 4
9. Blockchain & Space — 5 x 4

These tables imply a community of organizations/companies that participate in space-related activities. I will construct a master Organizations table and tables for associated projects like constellations, launchers, and upcoming spaceships **using a subset of the above tables**. This will allow me to probe the database for questions about the activities of various organizations, the relationships among organizations, trends in space activity over time, and so on.

2 Data Ingestion

I have copied the data from all of the above tables into individual sheets of an Excel document. I will use Excel to perform batch formatting operations to prepare the data for ingestion, just as I did for the Pokemon data in homework assignments. I will export the data from Excel into .CSV or .TSV files (comma- or tab-separated values, respectively). Then, using Vim, I will perform any necessary clean-up on the CSV or TSV data until they are ready for programmatic final parsing.

Next, I will use Python, Ruby, C#, or Lisp/Racket scripts to parse the CSV or TSV files. The scripts will parse the input files in order to construct `INSERT` queries send those queries to the DBMS using the appropriate database connector/API. After I create the tables, the scripts will populate the tables. In addition to populating the obvious base tables (Organizations, Constellations, etc.), the scripts will translate the implicit relationships in the CSV or TSV data into explicit SQL relationships in the form of junction table entries, foreign key values, and so on.

3 Questions

I will ask the following 20 questions:

1. How many different organizations have satellite constellations?
2. How many satellite constellations does each organization have?

3. What is the average number of constellations per organization, excluding organizations that have no constellations?
4. What organizations have both constellations and upcoming spaceships?
5. What organizations have both constellations and launchers?
6. What organizations have both launchers and upcoming spaceships?
7. What organizations have all 3?
8. How many constellations use more than 1 form factor?
9. What constellations use the most form factors?
10. What constellations use only 1 form factor?
11. How many projects (constellations, upcoming spaceships, and launchers) first launch each year? (List the number of launches in each year.)
12. What year saw the most launches (of any kind)?
13. The launchers listed are in many different statuses (stages of their lifecycles). How many launchers are in each stage/status?
14. Which launch type is most popular among launchers?
15. Which launch type is most popular among launchers that are currently in development?
16. What's the most expensive launcher? What's the least expensive? What's the average?
17. How much money has been spent so far on constellations and launchers (ignoring the ones for which we do not have funding numbers)?
18. How many constellations are full (all planned satellites are launched)?
19. How many constellations, launched satellites, and planned satellites pertain to each field?
20. What constellations have the most planned but unlaunched satellites?