**ETL Project**

**Project Team Three – Nicole Marcum, Hannah Lee, Mike Gleixner, Rick Gentile**



**Extravehicular Activity – Literally a World Away**

If you are lucky enough to look up and see a shiny light streaking across the night sky you may wonder, is that a bird? A plane ? No…. It’s neither. In all likelihood it is the International Space Station (ISS) a multi-nation effort serving as a microgravity and research laboratory since 1998.

As the International Space Station quickly vanishes into the horizon, you may ask;

Of those venturing 250 miles into the sky to visit the International Space Station which of these brave men and women have actually completed an Extravehicular Activity, or as we all know it a Spacewalk?

What about the pioneers of space travel in the 1960s and 1970s? Who were these individuals who left the protected confines of a perfectly fine spacecraft to venture into space tethered to their spacecraft only with a cable?

Our project examines the relationship between the visitors to space, the countries they came from, reasons for the Spacewalks and the number of Spacewalks for each person bolder to go where no person has gone before during the years of 1998 through 2017. We also look at the budgets of countries invested in exploring space and provide the background of Spacewalkers from the United States to show what the ‘right stuff’ is to be an astronaut.

We began with a deep dive, well a ‘shot in space’, for the data and information to build our database and utilized the following sources:



Our process of cleaning and aggregation was as follows:

1. We gathered CSV files detailing Space Walk Reason, ISS Space Walkers, US Astronauts, Space Agencies budgets and other countries budgets.
2. Cleaned the CSV files to incorporate the appropriate data field representation of the Spacewalker names by dropping middle names and middle initials and by adding the EVA number (Extravehicular Activity Number) to each mission when not present in the data field.
3. Using Jupyter Notebook, the incorporated process followed was:
   1. Imported the data from these CSVs into Dataframes using Pandas.
      1. Data/Austronauts.csv; Data/Space\_walks\_master.csv; Data/ISS\_spacewalkers.csv; Data/Space\_Agency\_Budgets.csv; Data/Spacewalkers\_with\_country.csv
   2. Transformed the DataFrames to ready for posting to the ISS\_Space\_Walkers\_DB
      1. Renamed the column headers
      2. Set indexes
   3. Confirmed the Database Connection
      1. Created the connection\_string
      2. Created the engine and confirmed the tables
   4. Loaded DataFrames into Database
      1. Austronauts\_master; space\_walk\_master; iss\_space\_walkers; other\_space\_agency\_budgets; space\_walkers\_with\_country; nasa\_budgets\_over\_time
   5. Merged then reviewed the data in the tables. (Tables Listed)
      1. Austronauts\_master
      2. Space\_walk\_master
      3. Iss\_space\_walkers
      4. Other\_space\_agency\_budgets
      5. Space\_walkers\_with\_country
4. Using SQL Alchemy, the data was load into SQL so that users may create tables detailing a variety of relationships: Spacewalkers per country, number of Spacewalks per year, Spacewalks as compared to countries budgets etc.
   1. Relational Database Decision - We chose to export our data to SQL, a relational production database. Since our data is text based it is aggregated easily with simple joins and was the most economical option.
5. Our final project provides a database containing tables with metrics detailing information on all individuals who have gone into space and sojourned out of their spacecraft in an extravehicular activity. We have included the missions that brought these men and women into space along with the budgets of those countries that have pushed the boundaries of the human spirit.
6. There is room for expansion in this database such as adding the experiments and research being conducted on the International Space Station. Also of interest would be to provide data on the relationships of the astronauts who have spent time on the International Space Station with their countryman and from other countries. Lastly information on the effects of time spent on the International Space Station on the astronaut in terms of physical and mental health.

Programming Overview Details



Database Iss\_Space\_Walkers\_DB visual schema.

