**ETL Project Proposal**

Levi-Evan Shantz-Mcmaniman

Zac Gouveia

Alexander Singur

Tom Callegari

The purpose of this project is to collect American financial data, data on terrorism, data on airline flights and combine them into a single database that is ready for queries and analysis of correlated factors between these multiple data sources.

Our plan to complete project:

* Find Data sources:
  + Need 1 main dataset and 2+ corollary datasets to clean and add into 1 database
  + What data will we use?
    - S&P 500 index data for last 7 years
    - Global terrorism data (needs to be cleaned to match dates with US S&P data)
    - Airline flight data (needs to be cleaned to match dates with US S&P)
  + What type of transformation/cleaning will we need to do?
    - Date group-by to match with S&P
  + What type of database makes sense for the data we are collecting?
    - Relational (PostgreSQL will do just fine)
  + Create tables in the database for our cleaned data
  + Insert the data
  + Right a query that will return joined tables in a usable format for analysis