**Transportation Through the Decade**

**Project Outline**

Matt and Rebecca wanted to get a glimpse into transportation trends for the decade between 2010 and 2020. With traffic growing, rail lines being extended, and air travel being more accessible, at least in our immediate community, out curiosity is aimed at growth of the use of the different transportation methods (air, rail, highway) nationally and the effects this has had on employment and personal spending within this field.

**Extraction**

One dataset was used from the public platform Kaggle. This dataset, US Monthly Transportation Statistics, is large and extensive covering monthly statistics from January 1947 through July 2020 on over 130 columns worth of information. This dataset was split into five smaller sets:

* Highway Vehicle Miles Traveled
* Personal Spending
* Rail
* Transportation Employment
* US Air.

The full dataset for US Monthly Transportation Statistics can be found here:[***https://www.kaggle.com/thaddeussegura/us-monthly-transportation-statistics***](https://www.kaggle.com/thaddeussegura/us-monthly-transportation-statistics)

**Transformation**

For each of the five datasets above, we did the following:

* Uploaded the CSV’s into Pandas Data Frames in Jupyter Notebook
* Deleted rows that did not contain information in at least one column (other than Index and Date)
* Renamed the columns so that they were in a more readable format
* Dropped unnecessary index rows
* Changed the data type of each column to the most appropriate type for the Load process
* Converted the Date column to data type datetime
* Deleted columns with dates older than January 2010

For the Passenger Rail CSV, we had an extra step of deleting the Passenger Rail Total Reports column as each row under that column had the same value and wasn’t adding any additional information to the dataset.

We also decided to split the US Air CSV into two smaller tables, one that included only Air Traffic information, and another that only included Air Carrier Cargo information. For each of those two smaller tables, we followed the above bullet points for cleaning.

**Highway Vehicle Miles Traveled Progress:**

**Before:Graphical user interface, text, application

Description automatically generated**

**Mid: A picture containing diagram

Description automatically generated**

**Final: A picture containing application

Description automatically generated**

**Passenger Rail:**

**Before:Graphical user interface

Description automatically generated**

**Mid:Graphical user interface

Description automatically generated**

**Final:Graphical user interface, application

Description automatically generated**

**Transportation Employment:**

**Before:Graphical user interface

Description automatically generated**

**Graphical user interface, text, application

Description automatically generated Mid:Graphical user interface, text, application

Description automatically generated**

**Graphical user interface, text

Description automatically generated Final:**

**Graphical user interface, text, application

Description automatically generated**

**Personal Spending on Transportation:**

**Before:A picture containing application

Description automatically generated**

**Mid:Table

Description automatically generated**

**Final:Table

Description automatically generated**

**US Air:**

**Graphical user interface, text

Description automatically generated**

**US Airline Traffic:**

**Before:A picture containing text

Description automatically generated**

**Mid:**

**Table

Description automatically generated**

**Final:**

**Table

Description automatically generated**

**Carrier Cargo:**

**Before:A picture containing graphical user interface

Description automatically generated**

**Mid:**

**Table

Description automatically generated**

**Final:**

**Table

Description automatically generated**

**Load**

We started the loading process by setting up our Postgres database. With the help of <https://app.quickdatabasediagrams.com/>, we created a table schema that was then exported to pgAdmin. All six of the tables that were created had the same primary key of “Date”. Since Transportation Employment had the most rows, the rest of the tables “Date” columns were foreign keys in relation to primary key “Date” in transportation\_employment table.

The next part in the loading process was connecting to the Postgres database from our jupyter notebook. Before our transformed DataFrames could be loaded, the column headers were made lower case to match our tables in pgAdmin. A sample query was then run to confirm the data was added successfully.

**ERD:**

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

Description automatically generated**

**Sample Queries:**

All Transportation vs. Personal Spending on Transportation Service

**Graphical user interface, text, email

Description automatically generated**

Passenger Rail Ridership vs. Rail Employment

**Table

Description automatically generated**

Personal Spending on Transportation Services vs. Highway Vehicle Miles Traveled All Systems

Graphical user interface, text, application

Description automatically generated