US Airline data from 1999 – 2009

Pseudo Code

Fly Date (Data Capture)

- ❖ Import Libs
  - ➢ Pandas
  - ➢ Matplotlib
  - ➢ Datetime
  - ➢ SciPy
  - ➢ Statsmodels
  - ➢ Numpy

- Load airline route data into Data Frame - define datatype for integers.
- Recast 'Fly Date' from YYYYMM format to separate columns of 'Month' and 'Year.'
- Create an empty list to store the chunks of 1,000,000 rows at a time
- Iterate over each chunk of data
  - ➢ Drop rows with null values
  - ➢ Recast 'Fly Date' from YYYYMM format to columns for 'Month' and 'Year
  - ➢ Append the modified chunk to the list
- Concatenate all of the chunks into one Data Frame
- Return the number of rows and columns in the Data Frame
- Reorder column order so 'Month' and 'Year' are repositioned to come after 'Fly Date
- Return updated Data Frame

Top 10 Route (Data Capture)

- Reorganizing and renaming Columns as a new Data Frame
  - ➢ Grabbing route and destination Data
  - ➢ Reorganizing passenger count by route
  - ➢ Merge appropriate columns with newly calc data
-

Data Cleanup

- Reordering and Naming columns
  - Reshape the Data
  - Remove all Rows without info or NaN info
  - Get number of unique rows in column (airport)
- Split date in to two columns
  - Convert Column to string
  - Assign new column to date info split by month and year

Calculate Percentage of seats used on Flight

- Access data Frame
  - Read Data Frame
  - Round math by 2 decimals places
  - Reordering Columns
  - Export data frame

Chart Generation

- Reads CSV Data
  - Count of unique routes, summing the value in the Flights column (number of flights in the given month)
  - Total number of unique routes
    - Get count of each occurrence of a destination code
    - Creates variables for Number of entries
    - stores 25 rows of data in new Data Frame (variable)
      - create a pie chart that shows the percentage of flights to each destination airport in the shortened Data Frame
      - Calc the sum of data in flights column
  - Group data in to one column and calc Sub of rows in 2 columns
- Creates variables for Number of entries
  - create a pie chart that shows the percentage of flights to each destination airport in the shortened Data Frame
  - Calc the sum of data in flights column

Percent Full vs Route Frequency

- Group data by 2 columns and calc the mean
- Merge data
- assigns it the length of the full Frame data frame after dropping the missing values.
    - Remove all nan Data
    - Assign it the length of the full frame after dropping missing values
- Create scatter plot