

Group #10

Members: Loren Williams, Naraendra Reddy Dendi, Tharani Kumaresan, Ankitha Yenna

Steps:

1. Created project and added group members
2. Uploaded csv file to BigQuery
3. Used DataPrep to clean the data
 - a. Started looking at columns with messy data that could be cleaned up(where are spaces, different data types, symbols, etc.)
 - b. Used DataPrep functions to clean the data in order to process in Notebook(screenshots of recipe process included below)
 - i. Changed column names
 - ii. Changed data types
 - iii. Removed white spaces
 - iv. Removed symbols
4. After data was cleansed and prepared, we ran it in DataPrep to ensure stored data was in BigQuery
5. To answer the questions for the assignment, we ran code in Python
6. A Data Dashboard was created in data studio showing the answers to five of the proposed questions. We created pie charts, scatter plots, and bar charts to visualize the answers to these questions.

Loren Williams updated today

- + Change CloseAmount type to String
- + Change DateReceived type to String
- + Change IncidentDate type to String
- + Change IncidentDate type to Datetime, mm-dd-yy hh:mm:ss, mm*dd*yyy*HH:MM
- + Remove whitespace from IncidentDate
- + Change IncidentDate type to String
- + Change IncidentDate type to Datetime, mm-dd-yy, mm*dd*yy
- + Remove whitespace from IncidentDate
- + Change ClaimNumber type to String
- + Trim whitespace from ClaimNumber
- + Trim whitespace from ClaimNumber
- + Remove whitespace from ClaimNumber
- + Change ClaimNumber type to String
- + Remove symbols from ClaimNumber
- + Remove whitespace from ClaimNumber
- + Remove whitespace from DateReceived
- + Remove whitespace from IncidentDate
- + Remove symbols from AirportCode
- + Remove whitespace from AirportCode
- + Remove symbols from AirportName
- + Remove whitespace from AirportName
- + Remove whitespace from AirportName
- + Remove symbols from AirlineName
- + Remove symbols from AirlineName
- + Remove symbols from ClaimType
- + Remove whitespace from ClaimType
- + Remove symbols from ClaimSite
- + Remove whitespace from ClaimSite
- + Remove whitespace from Item
- + Remove symbols from Item
- + Remove whitespace from ClaimAmount
- + Remove symbols from ClaimAmount
- + Remove symbols from ClaimAmount
- + Remove symbols from Status

- + Remove whitespace from Status
- + Remove whitespace from CloseAmount
- + Remove whitespace from Disposition
- + Remove symbols from CloseAmount
- + Change CloseAmount type to Decimal
- + Replace missing values in Disposition with the last valid value
- + Replace missing values in CloseAmount with the last valid value
- + Replace missing values in ClaimAmount with the last valid value
- + Replace missing values in Item with the last valid value
- + Replace missing values in Item with the last valid value
- + Replace missing values in ClaimType with the last valid value
- + Replace missing values in AirlineName with the last valid value
- + Replace missing values in AirportName with the last valid value
- + Replace missing values in AirportCode with the last valid value
- + Replace missing values in IncidentDate with the last valid value
- + Set IncidentDate to IFMISMATCHED(\$col, ['Datetime','mm-dd-yy hh:mm:ss','mm*dd*yyyy*HH:MM'], NULL())
- + Replace missing values in IncidentDate with the last valid value
- + Rename Close Amount to 'CloseAmount'
- + Rename Claim Amount to 'ClaimAmount'
- + Rename Claim Site to 'ClaimSite'
- + Rename Claim Site to 'ClaimSite'
- + Rename Airline Name to 'AirlineName'
- + Rename Airport Name to 'AirportName'
- + Rename Airport Code to 'AirportCode'
- + Rename Incident Date to 'IncidentDate'
- + Change ClaimNumber type to String
- + Rename Claim Number to 'ClaimNumber'
- + Rename Date Received to 'DateReceived'