# DAT 325 Project Two Template

# Executive Summary Report

Ashley Littles

## Data Set Anomalies

The first row is an example.

| **Key Value** | **Description of Anomaly** | **Plan for Resolution** |
| --- | --- | --- |
| *Example:*  *101339052 (Employee Number field)* | *Example: Last name missing* | *Example: Delete row* |
| Neutral, Positive, Negative and Unknown (airline\_sentiment) | The results come out to be alphabetic rather than numeric. | The results have to be replaced with the correct numeric value. |
| 5.67588E+17 (tweet\_id) | All the tweet ids are in scientific format and need to be switched over to number format. | Change format |
| American, Delta, Southwest, United, US Airways, Virgin America (airline) | The type of airline is shown as its full name rather than the shortened code. | Replace the airline names with their corresponding code. |
| Pacific time, Eastern time, Central time, Atlantic time, Sydney, etc. (user\_timezone) | There are parentheses and text within the parentheses. There are blank cells that are not set to N/A. | Remove parentheses and text within parentheses and fill the blank cells with N/A. |

## Data Types

You may need to add rows. The first three rows provide example answers. You will need to assess all header names from your chosen data set.

| **Header Name From File** | **Data Types Note** |
| --- | --- |
| *Example: State* | *Example: OK* |
| *Example: DOB* | *Example: Won’t Use* |
| *Example: Dates Employed* | *Example: Needs to be rounded* |
| tweet\_id | Format needs to be changed to numeric |
| airline\_sentiment | Format needs to be changed to numeric |
| airline\_sentiment\_confidence | Ok |
| negativereason | Fill the blank cells with N/A |
| negativereason\_confidence | Round up the values |
| airline | Replace names with code |
| airline\_sentiment\_gold | Ok |
| name | Ok |
| negativereason\_gold | Ok |
| retweet\_count | Ok |
| text | Ok |
| tweet\_coord | Ok |
| tweet\_created | Timestamp needs to be removed. |
| tweet\_location | Ok |
| user\_timezone | Parentheses need to be removed and any blank cells need to be set to N/A. |

## Specific Transformations Needed to Join the Data

You may need to add rows. The first row is an example. Responses may have one to three Excel functions.

| **Header Name From File** | **Excel Function One** | **Excel Function Two** | **Excel Function Three** |
| --- | --- | --- | --- |
| *Example: Smoker Status* | *Example:*  *IF(A1:A52=”Yes”,1,0)* |  |  |
| airline\_sentiment | IF(B2:506 = “positive”, 1) | IF(B2:506 = “negative”, -1) | IF(B2:506 = “neutral”, 0) |
| negativereason | Choose special select for blank cells. |  |  |
| airline | See lookup table IkupAirline | Find (command, F) Replace. |  |
| tweet\_created | Update format to Date. |  |  |
| user\_timezone | Choose select specific cells for the cells that contain parentheses. | Choose special select for the blank cells that need to be filled with N/A. |  |

## Executive Summary

In order to carry out the merger from Wayne Enterprise to Bruce, Inc we will need to clean up the errors within the database. To effectively clean the database from Wayne Enterprise we will use various functions within excel. There are many blank cells that need to be filled with the correct data and values that need to be rounded up in order to get rid of any decimals. A few columns within the dataset are not in the correct format so we will simply select the column and change the format. Overall, we are aiming for the data from the Wayne Enterprise database needs to meet the standards of Bruce, Inc.