Daniel Mejia

DSC 680 – Third Document

**Topic**:

For the third and final project, I will create a system which tells the user what the percentage of employee of their ethnicity are currently working.

**Business Problem**:

Companies have battle to create a company with multiple ethnicities and control the bad media and views from not having different ideas from different people. This could solve the problem by monitoring levels of employees and raising alerts for hiring, if necessary, in this company.

**Datasets**:

The data source will be from Kraggle. Only one file will be used Reveal\_EEO1\_for\_2016.csv, the file contains the required data to make the prediction required to get the percentage of employees for the year 2016. This will be a regression and not classification problem.

**Methods**:

Standard EDA will be used to explore and analyze the data. Also, a step further will be done by getting data important information to make recommendations. The data was explore using the provided function available to data frames such as describe, and similar methods. The standard deviation and mean will be measure for groups to understand anomalies and percentage above or below a threshold.

**Ethical Considerations**:

The model created could also cause damage to a company reputation if the levels are below the standard for the company. It could have the consequences of putting the company on a bad light, by pointing the fact that their hiring of Latino and other ethnicities is way below company’s standards.

**Challenges/Issue**s:

Almost all data is in categorical values and must be transformed to be used for the analysis. I will be doing counts of items depending on the fields such as gender, company, job, and similar fields. The categorical values must be counted, separated, and manipulated to find the answer required.

**References**:

The references will come from the documents obtain from Kraggle and the analysis obtained to get an answer. There will be another source to make the data more plausible with evidence, this will be a web article to be determine.