Group Name: **IDS\_GROUP052**

Assignment Name: **2019\_Cluster-DSE-IDS\_A1\_PS2**

Group Leader Name: **PARUPALLI SRINIVASA DINESH**

Group Member 1: **PARUPALLI SRINIVASA DINESH**

Group Member 2: **SWETHA MUTHYALA**

**Few Observations during EDA:**

* Dataset had few features which we are not required to answer the questions mentioned in the problem statement (which were part of the requirement), so they were removed.
* Job Category feature has only 2 null values, so we have not dropped them. Rest of the identified features were dropped. Refer to section with heading as “*Remove the simple features which are not really required for the analysis*"
* We renamed the feature names because after loading the dataset, while accessing the feature names we were facing with issues.
* We felt there is no need to perform the normalization on the dataset as per the questions asked in the assignment.
* Before proceeding with any of the data preprocessing, we verified with correlation among the features.
* We have noticed that there is an increased trend in the salaries over the years.
* Most of the salary come under Annual (91.3%) category.
* Job opportunities were in increasing trend over the years.
* Note: We have taken top 20 as to answer the question that have asked in the problem statement.

1. What are the highest paid Skills in the US market?

After the analysis on the dataset, we came to the conclusion that:

* Top 5 highest paid as per the Business Title are:
  + Deputy Commissioner, Public Information, M-VII
  + ADMINISTRATIVE MANAGEMENT AUDITOR
  + EVP for Real Estate
  + Deputy Commissioner, Bureau of Customer Services
  + Vice President Portfolio Planning, Project Development and Asset Mgt
* At ln#41, we have provided with the our analysis with regards to salaries vs no.of positions vs experience
* After analysis below are the topmost business titles come under the below job categories
  + Administration
  + Human Resources
  + Engineering
  + Architecture
  + Technology
  + Planning
  + Data & Innovation
* When applied the word cloud to the Job Category, then we came up with below topmost Job Categories:
  + Administration
  + Human Resources
  + Engineering
  + Architecture
  + Technology
  + Data & Innovation
* When applied the word cloud to the Business Title, then we came up with below topmost Business Title:
  + Director
  + Civil Engineer
  + Analyst
  + Project Manager
  + Coordinator
  + Deputy Director
  + Attorney

1. What are the job categories, which involve above mentioned niche skills?

After analysis on the data set, we came to the conclusion that:

* Architecture
* Engineering
* Technology
* Planning
* Research
* Inspections
* Legal

For more on the topmost job categories, please refer to ln#26 under Popular Job Categories section. We have created a custom logic to scan the topmost Job Categories.

1. Applying clustering concepts, please depict visually what are the different salary ranges based on job category and years of experience

Jupyter Notebook has all the necessary data visualization required for the job category, salary and years of experience.

We tried to make them readable and understandable as per our data analysis.