

### 1. Business Context

Using the given data set for New York City Current Job Posting data.

### 2. Business Problem Understanding

Focus on applying the learnt data analytics concepts and try to share your findings on following aspects:

- What are the highest paid Skills in the US market?
- What are the job categories, which involve above mentioned niche skills?
- Applying clustering concepts, please depict visually what are the different salary ranges based on job category and years of experience.



metadata.json



jobs.csv



data-d.xlsx

### 3. Expected Outcomes

The results should consists of

- The python script file or Jupyter notebook containing all the code for the proposed solution. Write all code in single file only, with proper comments. Don't include data file in the zipped file.
- A word document file containing answer to the following three sub questions (as asked above ) based on the analysis that you have carried out earlier.

#### 1. Step 1: Data Preparation – 10 marks

- Obtain a structure for the data using Python Programming Language – **1 marks**
- Create the required schema to read the data into the required format into rows and columns – **1 marks**
- Schema must be normalized, field types must be appropriate as per fields available. Proper data model e.g. – **8 marks**  
Select the appropriate features (columns) and parse the same, cleanup if required and convert to required categories

#### 2. Step 2: Identification of Variables – 10 marks

- Identify the required variables

#### 3. Step 3: Variable Selection – 10 marks

- Reason for the selection of the variable above

#### 4. Step 4: Feature Engineering – 10 marks

- What text parsing applied on *for the required fields*

**5. Step 5: Missing value or data – 10 marks**

- a. Missing values exist in following columns - 5 marks
- b. Special characters in some columns need to be handled - 5 marks

**6. Step 6: Analysis – 50 marks**

- a. What are the highest paid Skills in the US market? – 20 marks
  - i. Python code which queries on Top 10 Skills with Salary ranges – 15 marks
  - ii. If student have depicted using graphs, it would be good – 5 marks
- b. What are the job categories, which involve above mentioned niche skills? – 20 marks
  - i. Python code which queries and depicts Top 10 Job categories with above query result-set skills – 10 marks
  - ii. Graph to be plotted - 10 marks
- c. Applying clustering concepts, please depict visually what are the different salary ranges based on job category and years of experience. – 10 marks
  - i. Graphically plot all 3 dependent variables – i.e. job category, salary and years of experience
  - ii. Graph must be readable and understandable
  - iii. Graph type chosen
  - iv. Graph colour used
  - v. Legend and labels used