

Bits, pilani work integrated learning program



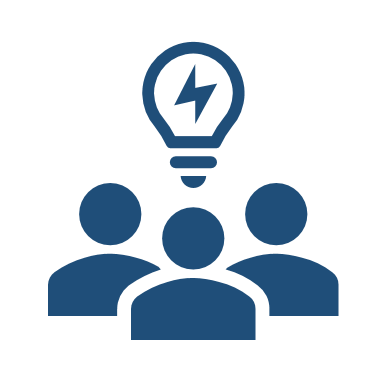
Introduction to Data Science

Analysis on New York Job Posting Data

Submission on 16 Dec 2019

**Introduction to Data Science**

**New York job posting data analysis (2019\_Cluster-DSE-IDS\_A1\_PS2)**



**Venkataramanan K | 2018AC04529**

**Balakavin Pon | 2018AC04531**

**Ponvani | 2018AC04559**

**Poornima J |2018AC04550**

Analysed and created in collaboration of

# Table of content

[Table of content 3](#_Toc27222713)

[1 | Problem statement 4](#_Toc27222714)

[Business Problem 4](#_Toc27222715)

[Data Exploration 4](#_Toc27222716)

[Feature Engineering 4](#_Toc27222717)

[Results 5](#_Toc27222718)

# 1 | Problem statement

## Business Problem

## 

Determine the below from the given data set for New York City Current Job Posting data.

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

## Data Exploration

The data is analyzed to summarize their main characteristics. Steps taken

* The percentage of null values are computed for each variable. If there are more than 30% null values, that variable is not considered for analysis.
* Handling missing values – Has been handled by replacing with most frequent occurring data (Mode strategy)

## Feature Engineering

Feature engineering is done to make the input dataset compatible with the machine learning algorithm requirements and improve the performance of machine learning models. Steps taken

* Remove unwanted characters such as special characters, unwanted whitespaces and punctuation.
* Remove Stop Words
* Extracting some specific attributes

## Analysis Done

* Highest paid skills was derived by the grouping by job category, preferred skills with salary range mean and picking the top 10 in this list.
* For clustering , bag of words transformer, word vectorizer was explored. K means clustering mechanism was used.

## Results

What are the highest paid skills in the US market?

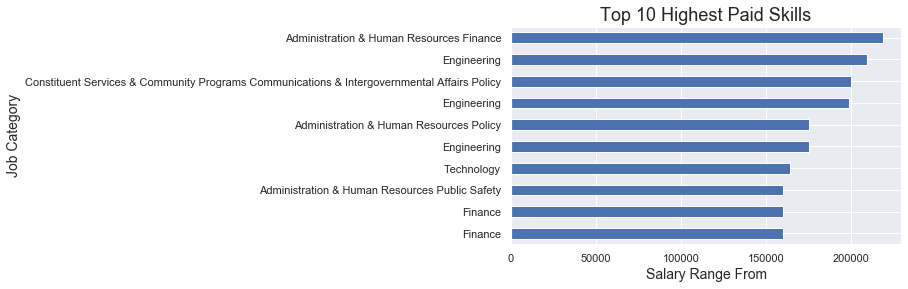
|  |  |  |
| --- | --- | --- |
| **Job Category** | **Preferred Skills** | **Salary Range** |
| **Administration & Human Resources Finance** | The Deputy Commissioner position requires exce... | 218587.0 |
| **Engineering** | The Deputy Commissioner position requires \te... | 209585.0 |
| **Constituent Services & Community Programs Communications & Intergovernmental Affairs Policy** | ERROR NAME | 200000.0 |
| **Engineering** | Extensive experience in wastewater operations ... | 198518.0 |
| **Administration & Human Resources Policy** | ERROR NAME | 175000.0 |
| **Engineering** | Required Skills \tIn depth knowledge of projec... | 175000.0 |
| **Technology** | The successful candidate will have years of s... | 164104.0 |
| **Administration & Human Resources Public Safety** | At least years of managerial experience in go... | 160000.0 |
| **Finance** | Clearly demonstrated experience trading money ... | 160000.0 |
| **Finance** | Expert knowledge of financial principles and c... | 160000.0 |

The highest paid skills in the US job market is :

The Deputy Commissioner position requires excellent leadership communication and client service skills as well as a passion for driving innovation continuous improvement and efficiency and implementing workforce strategies to recruit develop and retain the most qualified and diverse talent in a customer service environment The selected candidate will be expected to work closely with Agency and oversight counterparts and must have excellent management skills focused on staff alignment with the Bureaus vision goals and career development In addition he/she must be an effective team player capable of sustaining a culture of excellence in the delivery of customer service to internal partners and possess the ability to establish and maintain effective external partnerships and collaboration. Specifically the following skills and abilities are required to be successful in this role years relevant work experience leading customer focused functions of which at least years must have been in a senior leadership position BABS degree MBA ,advanced degree preferred .Strong analytical skills with ability to interpret data and trends diagnose problems and implement action plans to resolve issues outstanding written and oral communication skills. Comfort in managing concurrent projects in a fast paced results driven environment with equal enthusiasm for high level strategic planning and tactical daily execution Indepth knowledge of Federal City and State governmental rules and regulations related to the position NYC experience is a plus Proficiency in Microsoft Office Suite

The Job Category is- Administration & Human Resources Finance, Accounting, & Procurement Building Operations & Maintenance

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.

