

Wuzzuf Data Exploration (2014-2015)

1. Introduction

There are 2 files of data: Jobs Posts and Applications Samples (2014-2015). After checking the 2 files, I have found a messy columns, like City column, Experience Years, Job Description and Job Requirement. So, I will gather and clean the data to get perfect Insights, then save edited data to new csv files.

I have used jupyter to clean the data and you can find the code in this project in the github repository.

2. Data Cleaning

It will be 3 phases of data cleaning:

- First, fix the missing data in Currency and Payment Period Columns. I filled the missing data in Currency Column with 'Egyptian Pound' and for Payment Column with 'Per Month'
- Second, find the Duplicates data.
- Finally, Check data type of Columns and I have found that post_date in Jobs Posts data and app_date in Applications data are strings I converted them to datetime.
- City Column also needs cleaning, I thought a lot to find the best solution.
Finally, I decide to collect cities in groups like: Cairo_gov, Giza, Delta, Abroad, Sinai, 10th of Ramadan city, 6th October City, Upper Egypt, Red Sea, Canal, West Egypt, All Egypt, other and Remote.
- Experience year, I have fixed it with regex, find all numbers and make new ranges: 0, 0-3, 0-5, 5-7, 8-10, 10-15, more than 15.
- Job Description and Job Requirement columns includes html tags, so want to only show the content without tags. So I will strip html from string using HTML parser.

3. Analyzing Data

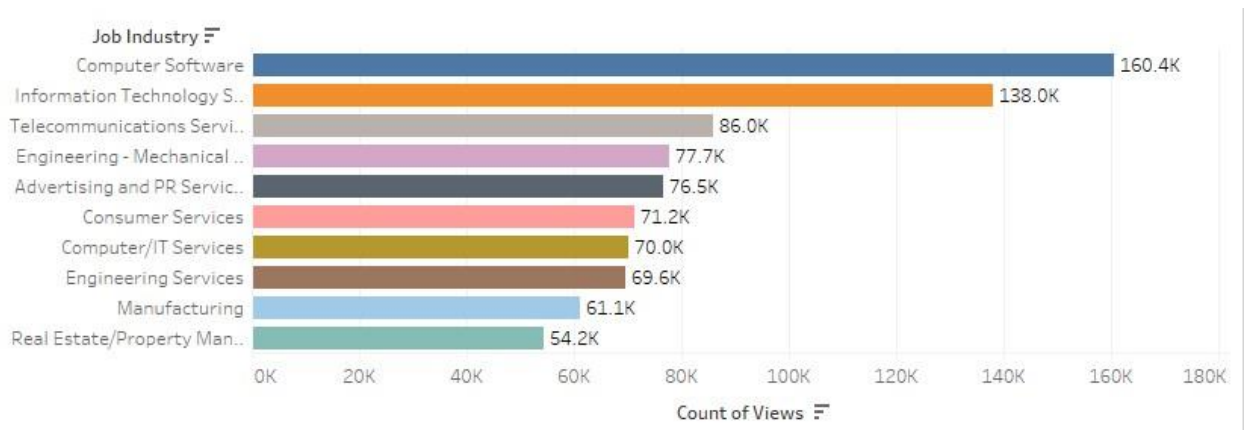
Here's the forecast and trend line for Count of vacancies to know if the vacancies in the next years will increase or not.

So, Trend line details is p-value: 0.0001 it's less than 0.05

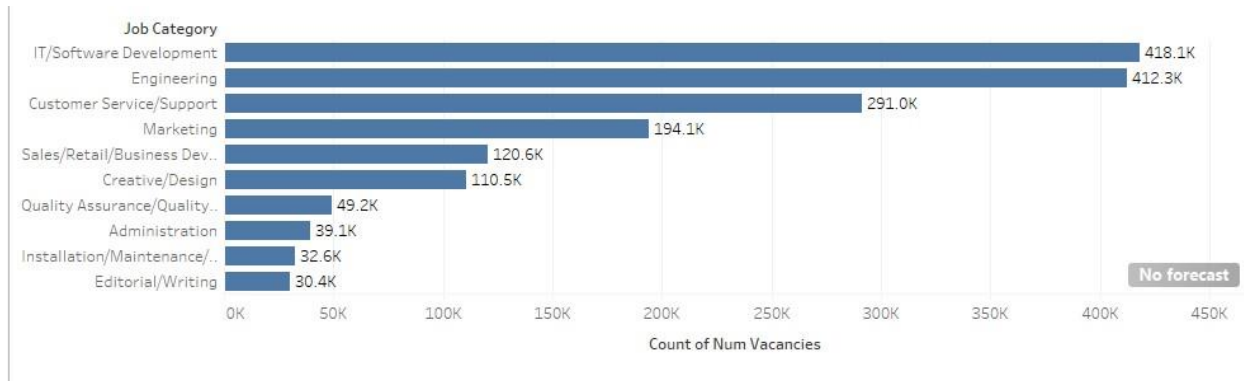
And R squared:0.7, so it's good for both we can depend on this values.



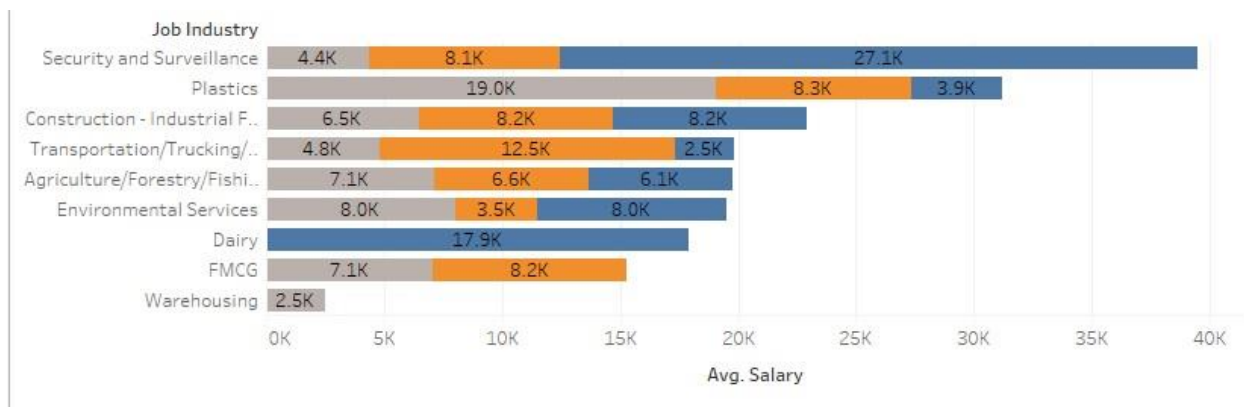
Here the Job Industry and count of views, here we could see the most job industries that has views: Computer Software.



The top categories that has the most vacancies: It/Software Development



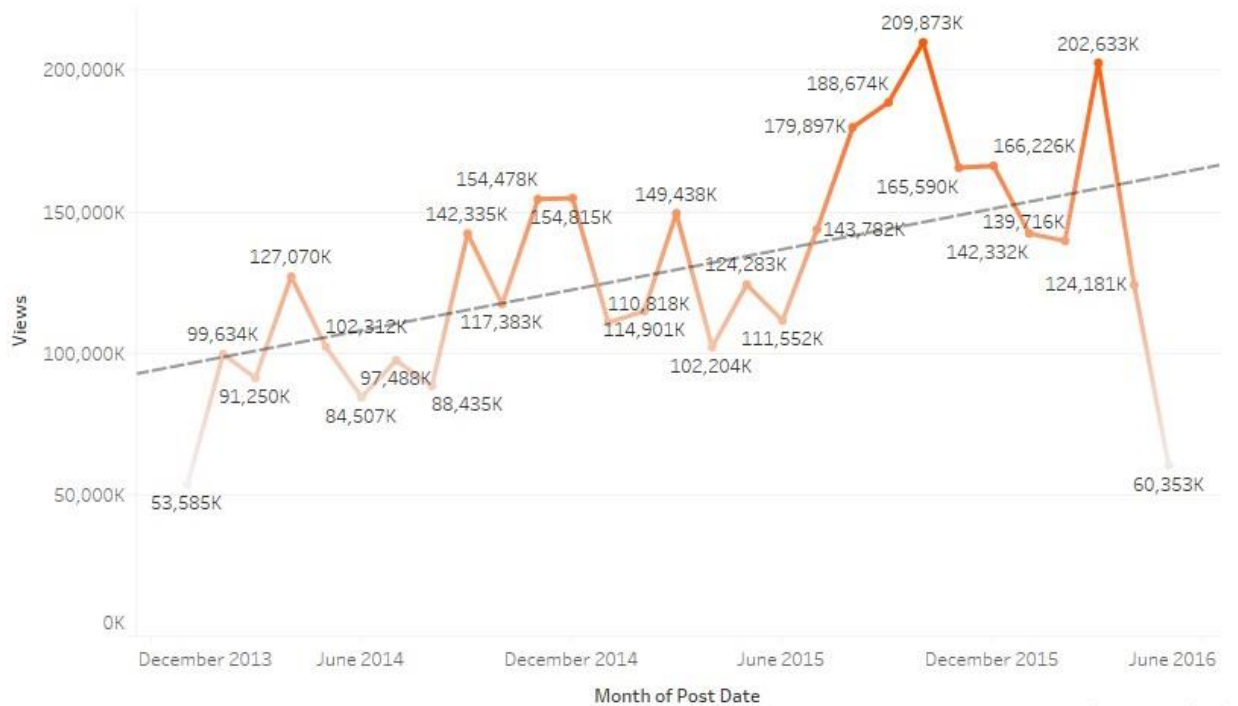
Job Industry Avg Salary per Years



Trend line for the views per months

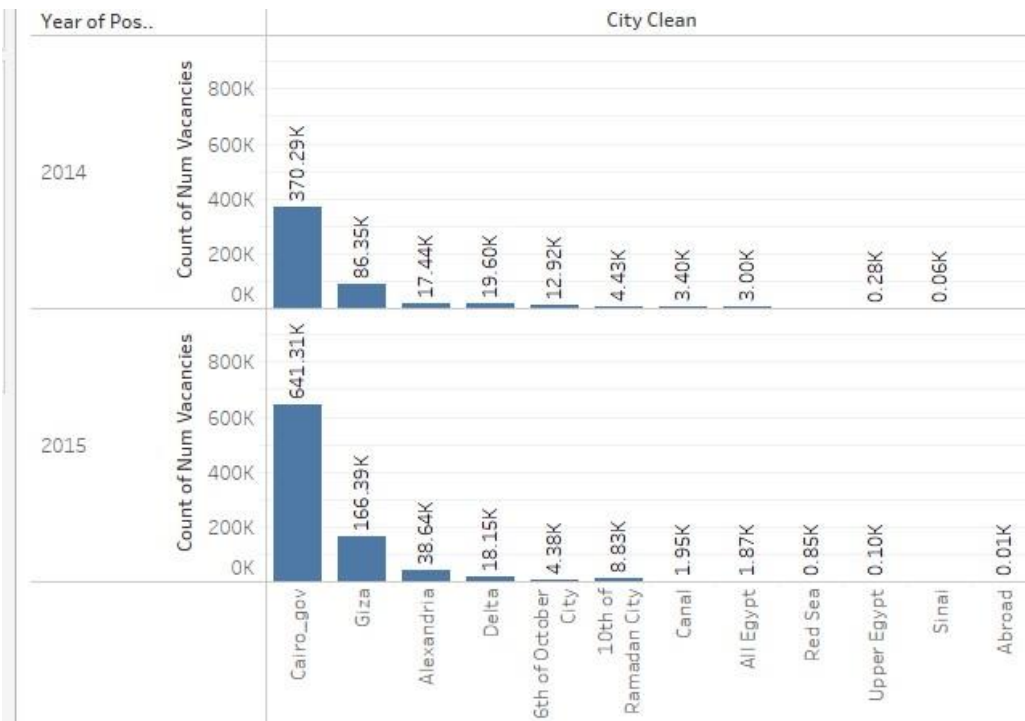
p-value: 0.003

r squared: 0.29



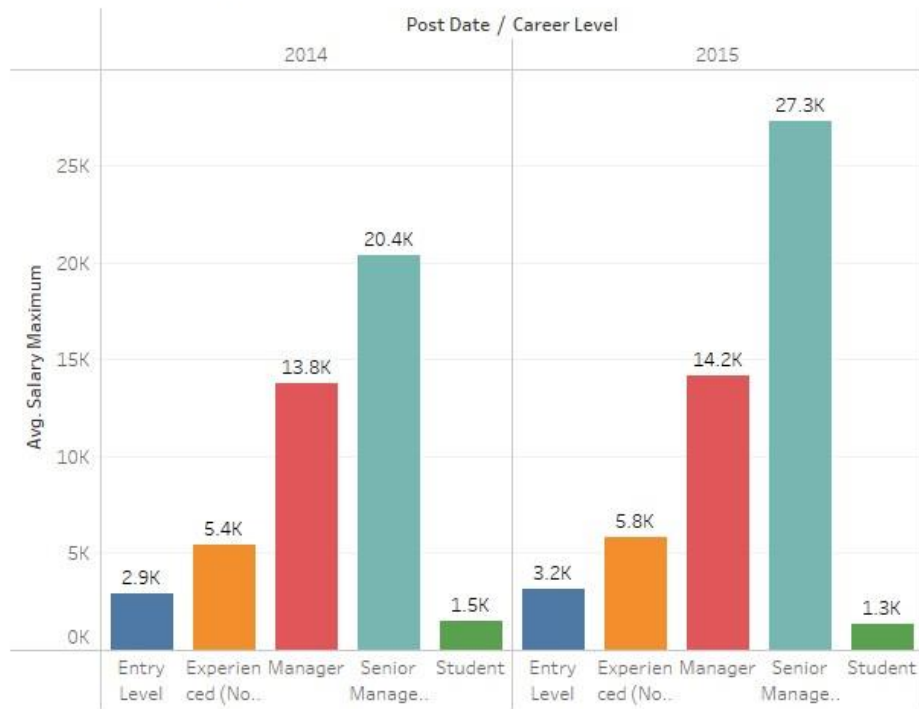
Count of Vacancies for cities per 2014-2015

We saw that highest number of vacancies was in Cairo then Giza



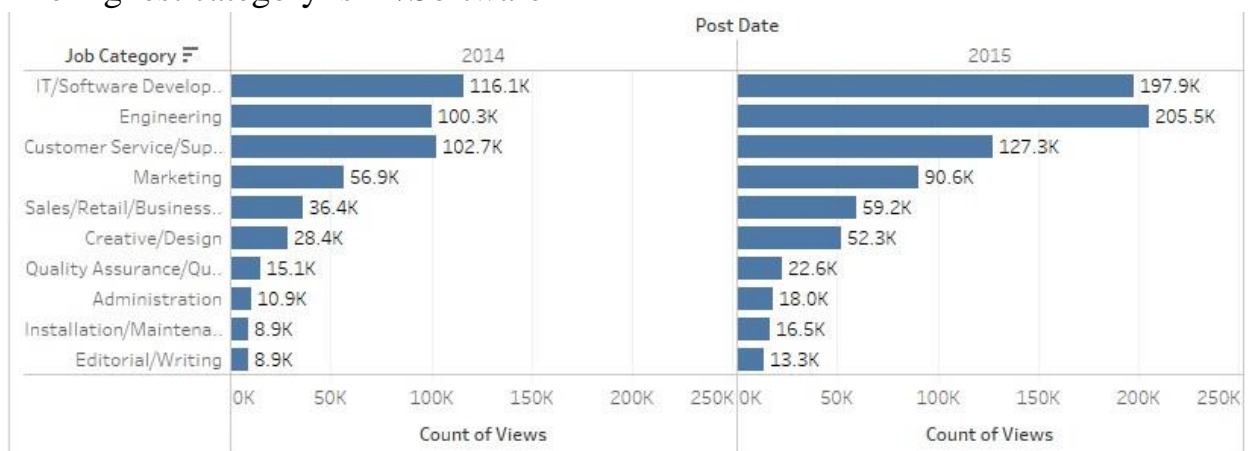
Avg Salary for Career level
the highest level is Senior in 2014-2015

Career level Avg Salary

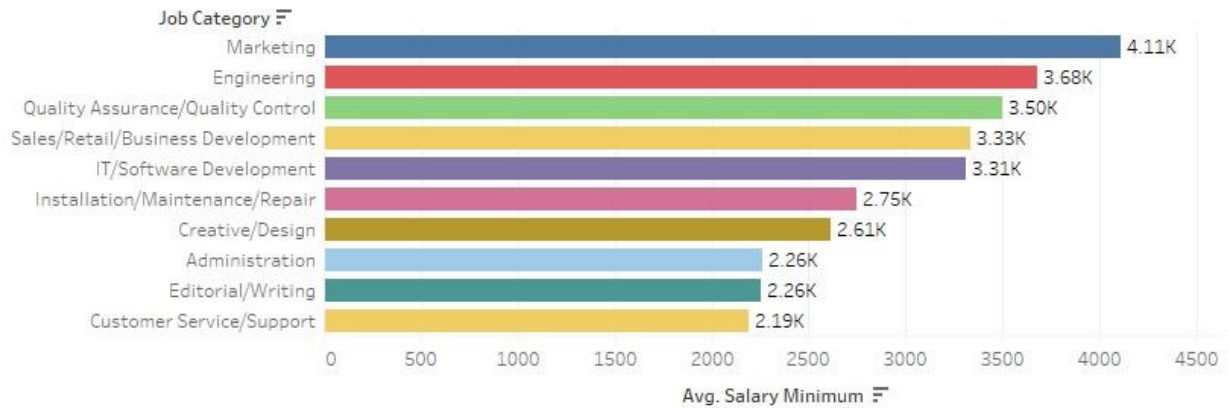


Count of Views of job category, we can see that views increased from 2014 to 2015

The highest category is IT/Software



Avg Min salary of Job Category: Marketing



Avg Max salary of Job Category: Marketing



4. Summary

- Number of Vacancies will increase for the next years will increase or not.
- The top categories that has the most vacancies: It/Software Development
- The Top cities of Vacancies are Cairo, Giza and Alex.
- Avg Salary for Career level, the highest level is Senior in 2014-2015
- Avg Max and Min salary of Job Category: Marketing