Resume Extraction project EDA.

Nada Belaidi

6/08/2021

Agenda

Executive Summary

Problem Statement

Approach

EDA

EDA summary and recommendations

Model recommendations

Executive summary:

The Client:
HR departements.

Data Set:
A json file containing 200 resumes.

Objective:
Providing meaningful insights through analyzing the data provided in order to help the client's HR departement enhance their performance.

Analysis plan:
-Graduation year analysis.
-Candidates location analysis.
-Companies candidates worked at analysis.
-Condidates degrees and universities studied at analysis.

Recommendation:
I recommend that the HR department takes more care of the job description and the requirements provided in order to have a more accurate candidates resumes.

problem statement :

Problem:

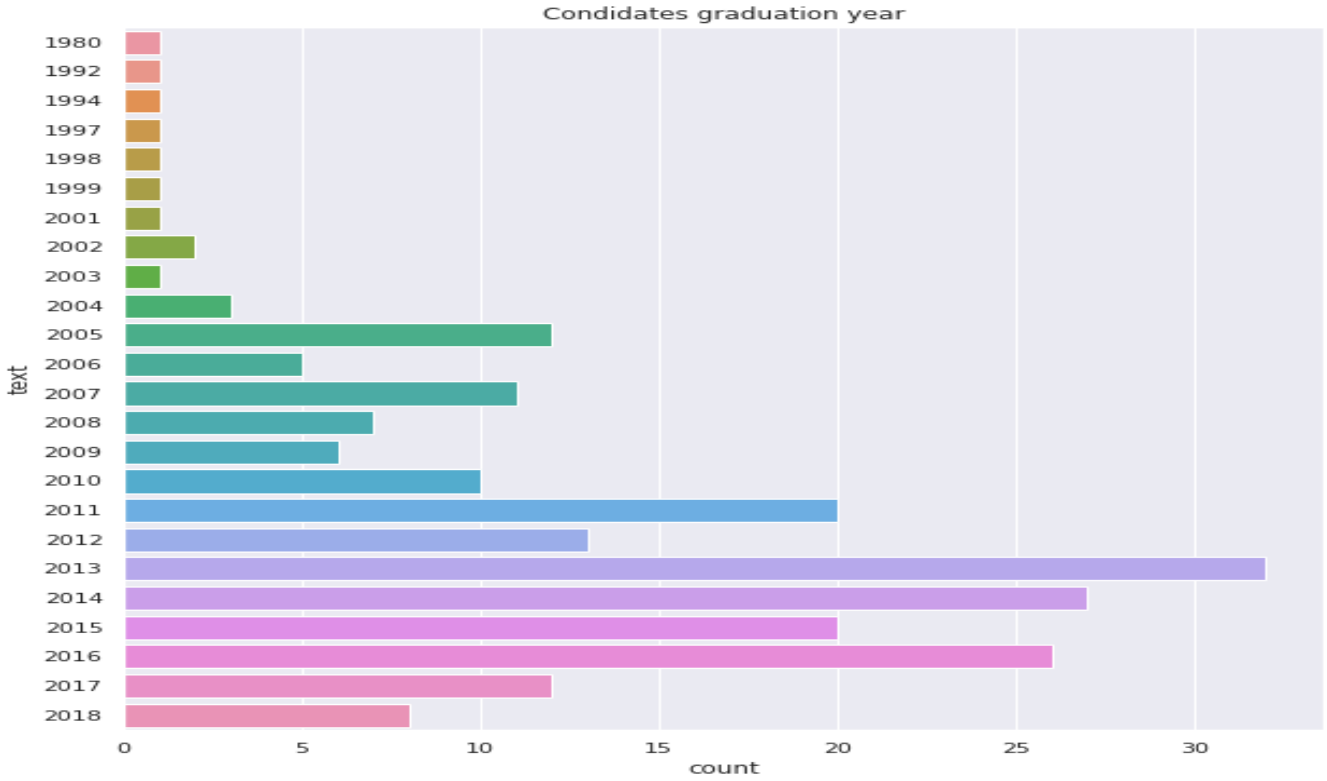
Resumes contain surfeit information that is not relevant for the HR/authority, and they have to manually process the resumes to shortlist the promising candidates for them. This EDA is dedicated to them to help them enhance their performance.

Approach:

Prescriptive analysis is the frontier of data analysis, combining the insight from all previous analyses to determine the course of action to take in a current problem or decision. And that is exactly why it was my choice for this study case.

EDA

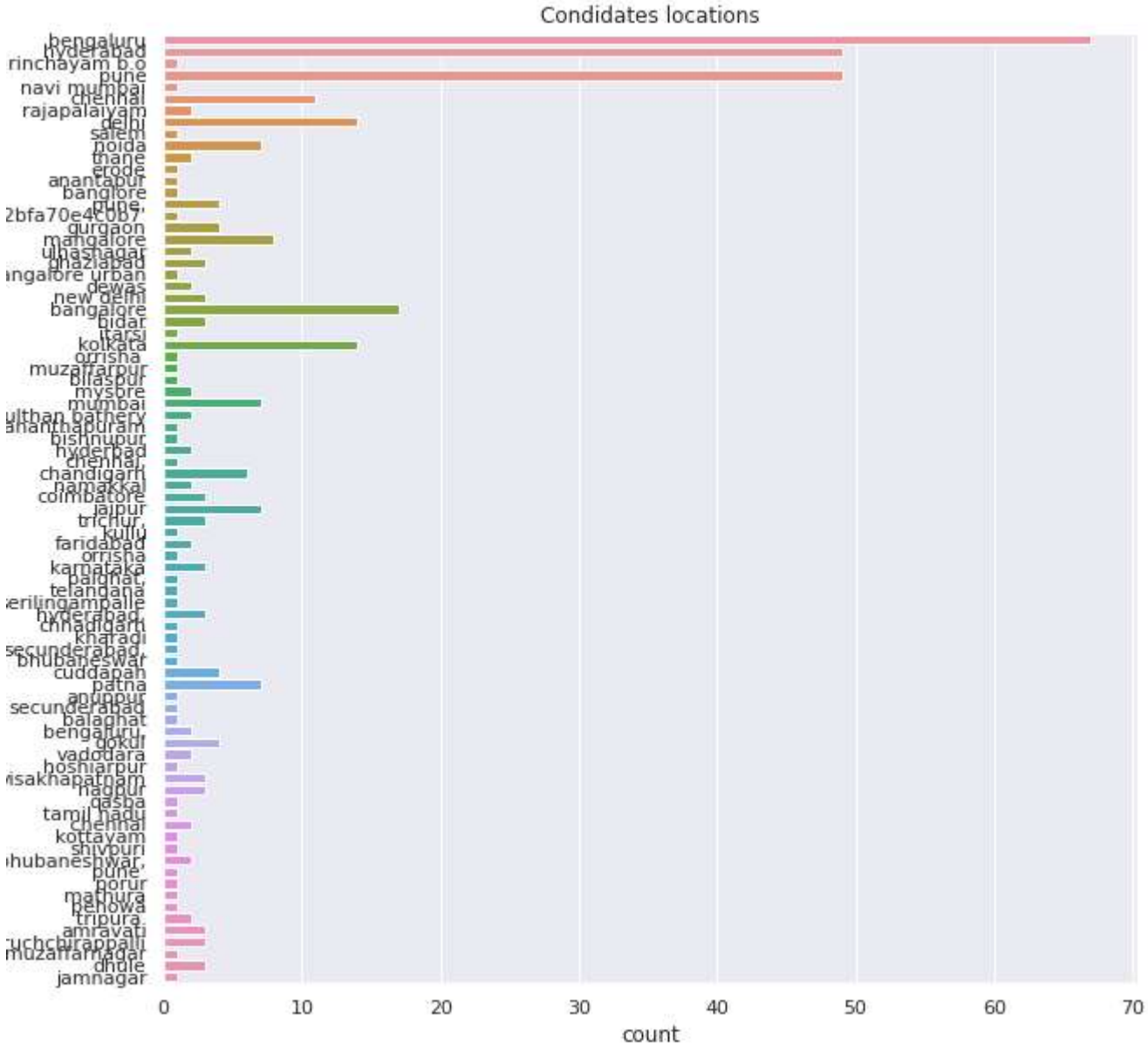
Graduation year analysis:



We can see that most of the candidates graduated after 2005.
16% of the candidates graduated on 2013 only and 50% of hem graduated between 2016 and 2013.
No condidates graduated after 2018.

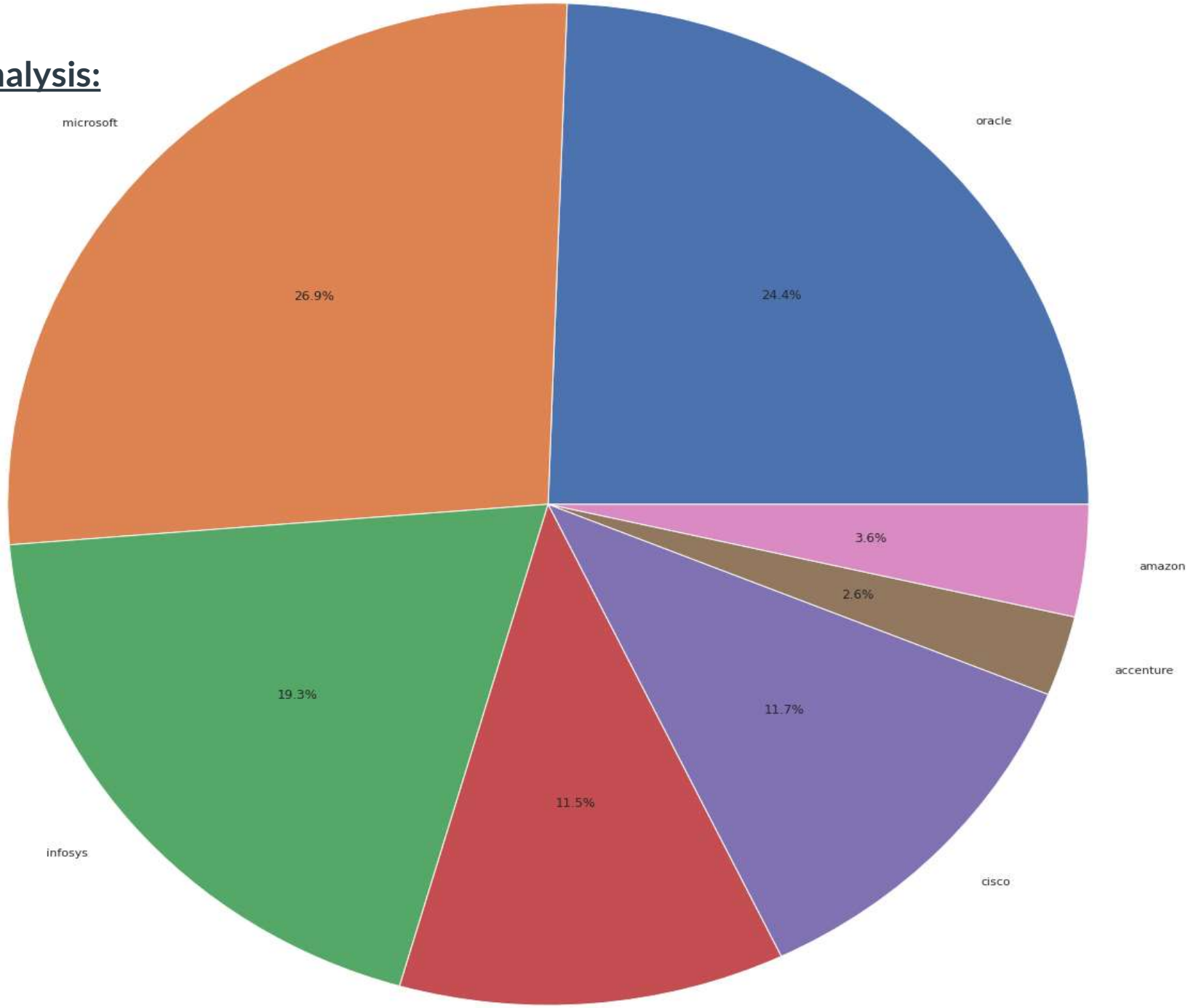
Candidates location analysis:

The majority of the candidates are located in Bengaluru (67 out of 200), Hyderabad and Pune (49 in each one).



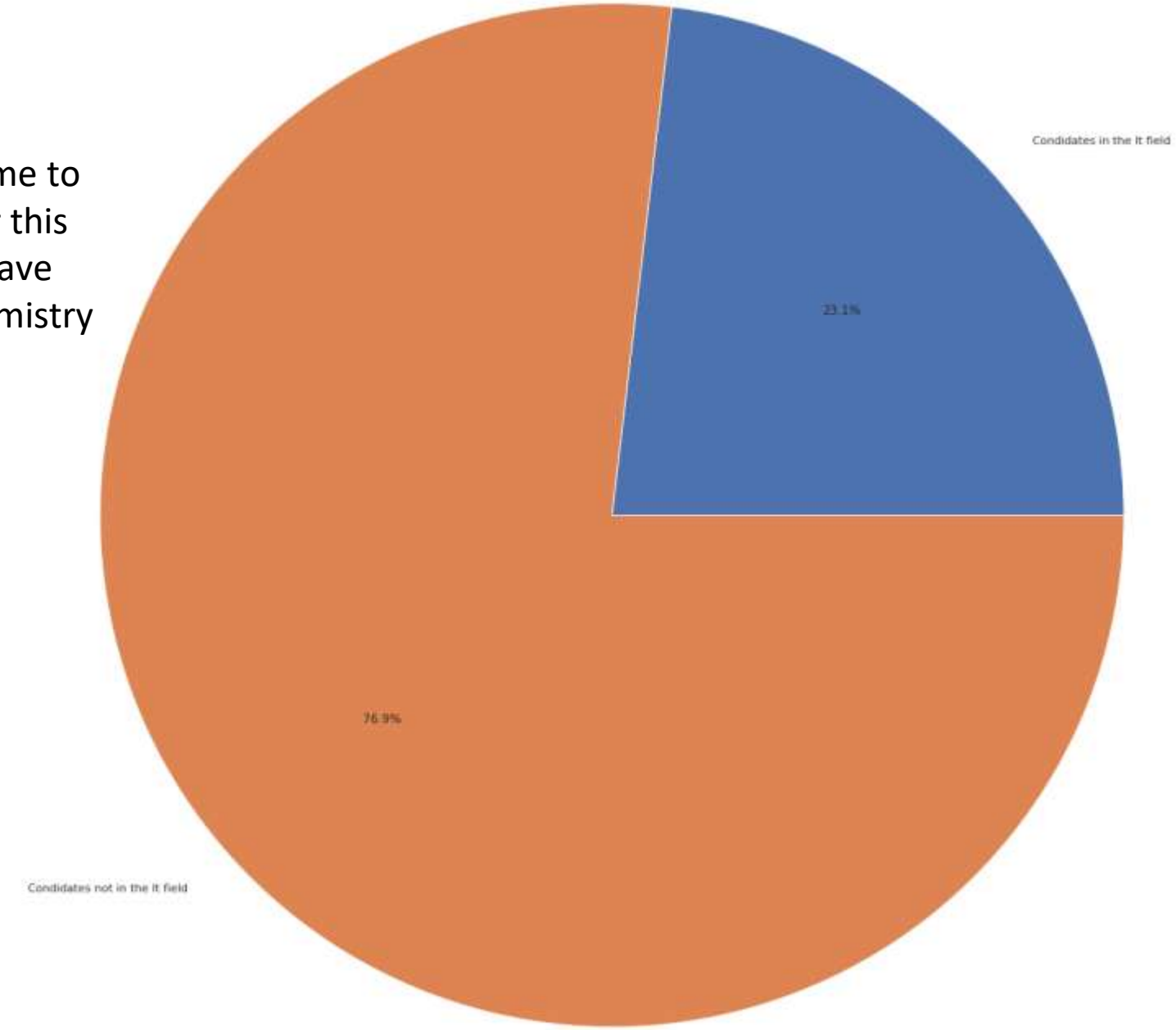
Companies candidates worked at analysis:

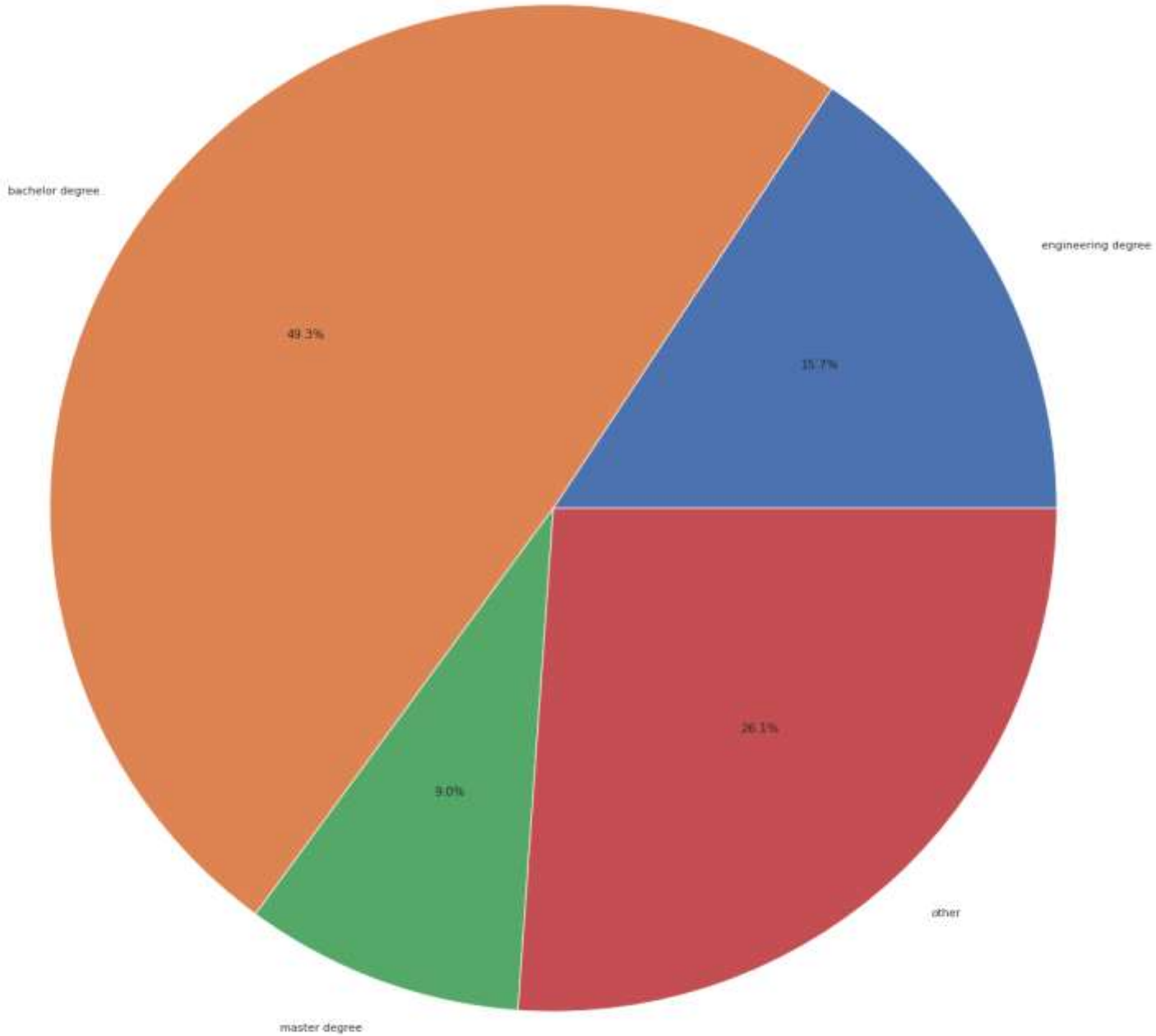
As for, the ‘Companies worked at’ data frame the following chart shows that Oracle, Microsoft, SAP, Amazon,Cisco, NIIT and Infosys are the ones with the biggest numbers of candidates having worked at them in the past.
At least 155 out of 200 (26.9% as we can see through this chart below) candidates worked at Microsoft before, 142 out of 200(24.4%) worked at Oracle.



Candidates degrees analysis:

-Analyzing the “Degree “data frame allowed me to find out that only 23.1% of the candidates for this job are in the IT field while the rest of them have different fields of study such as: business chemistry electronics etc. .





49.3% of the candidates have a bachelor degree.
15.7% of them are engineers.
9% candidates have a master degree.

Universities candidates studied at analysis:

Analyzing the universities, the candidates studied at I figured out that almost each one went to a different college.

EDA summary and recommendations:

-The EDA performed on our data has shown that the candidates who have applied are quite different, they don’t share to the same field of study nor the same degrees. They went to different colleges and they don’t have the same skills.
-Depending on these insights, the client’s HR department can request the elimination of some candidates’ categories depending on the job profile needed (example: the client needs candidates with engineering degree)
-Also, this EDA has shown that those who applied for this job are quite different especially when talking about the fields of study, I recommend that the HR department takes more care of the job description and the requirements provided in order to have a more accurate candidates resumes.

Model recommendations:

Depending on the data we have , our business and data science objectives I suggest the following models to work with in order to build the resume extraction app :

BERT:

Bidirectional Encoder Representations from Transformers (BERT) is a transformer-based machine learning technique for natural language processing (NLP) pre-training developed by Google.

BI-LSTM:

A Bidirectional LSTM, is a sequence processing model that consists of two LSTMs: one taking the input in a forward direction, and the other in a backwards direction.

Thank You