Aflore Data Science Challenge

Thank you for your interest in Aflore! We appreciate the time you spend on this challenge and are happy to answer questions.

1 Overview

The purpose of the challenge is for you to demonstrate your data science skills in the context of a semi "real-world" problem.

Please use the attached Jupyter Notebook (*aflorepyter.ipynb*) to do the challenge.

Using the Jupyter Notebook with Python 3 makes it easier for us to assess your work as is the language we use here every day.

We recommend to use the few libraries in the requirement.txt file, but feel free to use whatever library you want, just please add them to the attached requirement file.

We are more than happy to answer questions and point you to some good learning materials if you are new to Python.

2 Challenge

The challenge consists of Two steps:

- Analyzing the attached data and answering to the questions below (also present in the Jupyter Notebook).
- Make a further EDA (Exploratory Data Analysis) to provide any insightful result.

Keep in mind that this is a challenge to evaluate your level so, even though you are not able to answer all the questions, feel confident to send us all the results you can get.

The purpose is to evaluate your skills so **argue and explain your answers**.

Feel free and more than welcome to play around with the data and give any result you like, but try first to answer as many questions as you can.

Note. Answer either in Spanish or in English, whatever of those two languages you feel more confortable.

The answer should be submitted within three days from the day you received the challenge.

2.1 Data

Please, download the data. The data are available as csv files and sqlite db *aflore_sqlite.db* that is already loaded in the Jupyter Notebook.

The dataset consists of a toy dataset with four tables. Each table with its fields is described below:

candidate:

- id: candidate id.
- location: residence city of the candidate.
- birthdate: candidate's birthday date.
- gender: gender of the candidate.
- salary range: salary range of the candidate in the last job.
- transport: type of transport the candidate has.
- created on: date the candidate signed up in the platform.

work experience:

- id: work experience id.
- candidate id: candidate id.
- job title: job title of the experience.
- industry name: industry of the work experience.
- location: city of the work experience.
- sale type: Type of sale of the work experience.
- personin charge: number of people in charge the candidate had in the work experience.
- starting date: starting _date of the work experience.
- ending date: ending date of the work experience.
- observed: True if the work experience has terminated False otherwise.

job listing:

- id: job listing id.
- business id: id of the company that published the job.
- title: job title of the job offer.
- published on: job offer publication date.
- location: city of the job offer.
- salary range: salary range of the job offer.
- industry name: industry sector of the company publishing the job.

candidate application:

- candidate id: candidate id.
- job listing id: job listing id.
- date recommanded: date on which an invitation to apply to a job offer was sent (generally there's a small difference of time between date recommanded and published on in job listing table).
- date accepted: date on which a candidate applied to an invitation received and hence recommended to the job.
- date read: date on which a company read the profile of the candidate that applied to an invitation received (it can be not null only if date accepted is not null).
- date liked: date on which a company likes the profile of the candidate that applied to an invitation received (it can be not null only if date accepted is not null).
- date disliked: date on which a company dislikes the profile of the candidate that applied to an invitation received (it can be not null only if date accepted is not null).

2.2 Question to answer

- Provide the total number of candidates that have received an invitation and the total number of candidates that have accepted an invitation.
- What is the total rate of accepted invitations? What percentage of candidates have accepted an invitation?
- What is the distribution of candidates by age? What is the distribution of candidates by gender? What is the distribution of candidates by salary range?
- Consider the distinct segments age, gender, salary range and candidate residence city (only the first 10 most populated cities in the db). Are there significant differences of acceptance rate within such segments? Argue the answer.
- What is the average work experience time?
- How many work experiences a candidate have on average? What is the distribution by industry of work experience?
- What is the average number of candidates that have accepted an invitation per post for the jobs published in the month of January 2019? Plot the cohort time line (on job published) by month.
- What is the average number of candidates read, liked and disliked per post and company?
- Plot the timeline of total jobs published grouped by month, has there been a growth? What is the average monthly growth rate?
- Assume the job growth rate stays constant to the average monthly growth rate, what would be the expected number of jobs published after one year (January 2020)?
- Can you try to predict what would be the total number of jobs published in the first week of April 2019?
- Given that a candidate has created his account between 2018-01-01 and 2019-01-01 and has received invitations within 30 days from the account creation, what's the percentage of candidates that accept at least one?