ANALYSIS AND VISUALISATION OF DATA SCIENCE JOB OPPORTUNITIES

Giada Palma, VR471280, giada.palma@studenti.univr.it

Master's degree in Data Science, University of Verona

ABSTRACT

The data scientist is a relatively new key player in organizations. They are part mathematicians, part computer scientists, and they rule the world of big data.

Businesses today are wrestling with volumes of unstructured information that's a virtual gold mine, which can help boost revenue when unearthed. But they really need professionals who can dig in and find valuable business insights.

That is what the data scientist does; that is why they are highly sought after and well paid.

INTRODUCTION

The project will focus on the analysis of the job opportunities for data scientists, as I am a student in the data science master's degree, I am very interested in deepening my knowledge about the offers and what is expected from such a role.

The project will start with a classical EDA process enriched with some graphical visualisations of the most important statistics that will emerge.

Then it will follow the application of a natural language processing algorithm to discover word frequencies in descriptions of the jobs offered to data scientists

In conclusion the most important information will be shown in a dashboard.

The whole project will be implemented using jupyter notebooks and standard libraries implemented in python.

For the presentation of the project, I will implement a web application using streamlit python library.

SCENARIO OF USE

I have decided to implement the project using a web app, because we can suppose that the dashboard might be used by universities to show future students what are the positions offered for data science graduates, or might be used by agencies, helping people to find jobs, to show where are data scientist more requested, or might be simply used by people looking for jobs or interested in the data scientist figure.

A web application is easily accessible from any device and location, and with the help of a dashboards it is easy to have a quick first glance on the most important information.

DATA AND TASK ABSTRACTION

WHAT

- DATASET TYPE:
  a table contained in a comma separated file (GitHub - picklesueat/data_jobs_data).
- DATA TYPES:
  - categorical attributes, like job title and company name.
  - quantitative attributes, like salary, number of employees, company ratings.

WHY

- ACTIONS:
  - Extract and clean data
  - Analyse data to understand the content
  - Create visualisation to present results
- TARGETS:
  - Salary Estimate
  - Job Description
  - Company Name and Rating
  - Location of the Head quarters

HOW

- MARKS: lines, points, areas
- CHANNELS: position, size, and luminance for ordered attributes and spatial regions, shape and colour hue for categorical attributes.

SOLUTION AND PROPOSED IMPLEMENTATION

As a first step, I will start my project by performing some exploration and cleaning on the dataset.

Secondly, I will show some basic statistics on the data.

Thirdly, I will create some graphical visualisations to show data based on interesting queries.

Moreover, I will perform natural language processing on the job descriptions.

Lastly, I will create a dashboard composed by the more interesting KPIs found during the analysis.