**Big Data Project**

Average salary Intern/Mid/Senior/Manager in different industries

*Marcin Pestka, Jakub Prach*

The goal of the project was to gather and analyse data about current job offers posted on the RocketJobs platform grouped by industry (HR, Sales, Engineer, Marketing, Finance) and level (Intern, Mid, Senior, Manager). The results of our projects could be used in such cases:

* High school students can decide what industry is most profitable at the moment and choose their future studies based on the results of the project.
* People employed in one of the industries can compare their salaries to the average in the job market. It could convince them to ask for a raise or start looking for a new job.

**Type of data**

The type of data collected in this project is secondary and process-mediated.

**Data source**: <https://rocketjobs.pl>

**Quality of data**

**Business environment** – Salaries in the industry can vary depending on the economic situation.

**Privacy and security** - Data collection is allowed.

**Complexity** - The data has clear categories and thus is easy to analyse.

**Completeness** – Data is complete.

**Usability** - Data can be processed quickly and without much difficulty regarding its size.

**Time - related aspects** -When the website updates, the data still can be collected in the same way without the need for waiting**.**

**Representativeness** - The data only represents 5 industries.

**Linkability** - Data can be easily linked with other sets provided that it has similar format.

**Coherence** – The average salary for an industry changes over time as the need for an employee with particular skills can change.

**Project implementation**

**Methods**

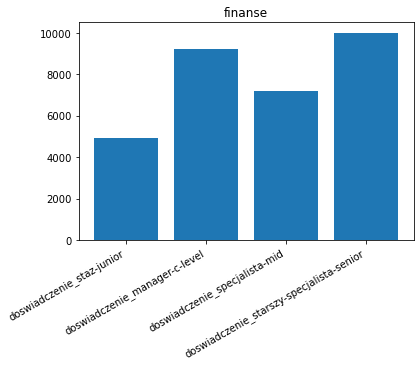
To gather the data from the RocketJobs platform we implemented a web scraping approach using **python**, **requests**, **BeautifulSoup** and **matplotlib** to present data visually.

Example data of sales job posts:

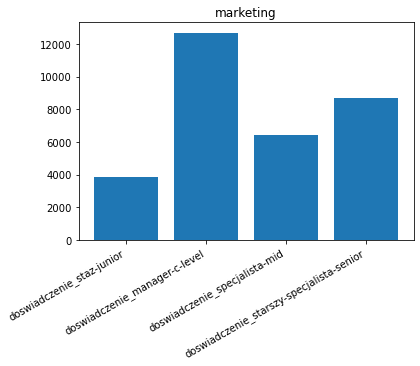
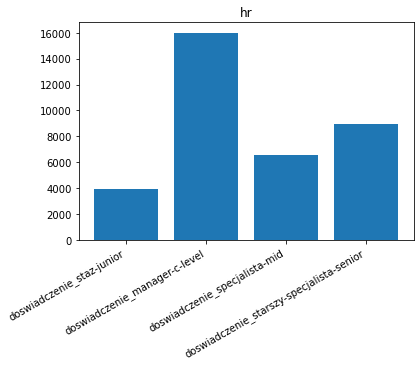
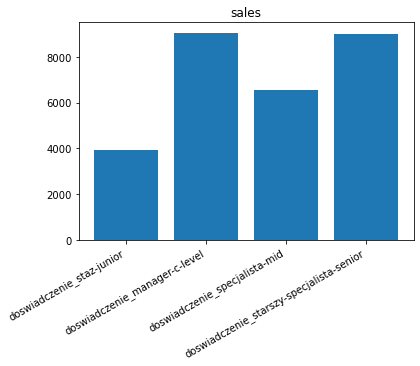
Obraz zawierający tekst

Opis wygenerowany automatycznie

**Results**

X – Experience level

Y – Average salary

****

