

UBC Salaries: Exploratory Analysis of Gender

Jade Bouchard

Table of contents

Aim	1
Data	1
Methods	1
Results	2
References	2

Aim

This document explores The University of British Columbia (UBC) faculty salaries based on guessed gender.

Data

Salary data was sourced from (University of British Columbia 2020). To access individual financial reports, click on the yearly links under the header **Statement of Financial Information (SOFI)**.

Gender data was inferred using first names of staff members. In order to guess gender, I used baby name datasets Sharma (n.d.).

Methods

The Python programming language (Van Rossum and Drake 2009) was used to perform this analysis.

Data collection

As mentioned earlier, for UBC salary data, I used the PDF salary information that UBC releases every year (University of British Columbia 2020). The data collection code performs the following tasks:

- Loads salary data that has already been collected as a dictionary from a pickle file. The
- Uses `requests` to access the UBC Financial Reports webpage.
- Extracts links to annual salary PDFs by locating the "Statement of Financial Information (SFI)" link.
- Uses the link url to map the fiscal year to the corresponding report link.
- If there are any links (and associated PDFs) for which we have not already collected salary data, it is added to the dictionary.
- Any new data is added to the salary data dictionary and saved as a pickle file.

Data cleaning

- Clean and Transform Data:
 - Removes unnecessary content
 - Uses regex to process the raw text data into a structured DataFrame
 - Splits names into first and last names.
 - Converts salary values to a numeric data type.

Results

References

- "Kaggle." n.d. <https://www.kaggle.com/datasets/kaggle/us-baby-names/data>.
- Sharma, Anany. n.d. "US Baby Names." <https://www.kaggle.com/datasets/ananysharma/indian-names-dataset>.
- Statistics Canada. n.d. "Table 17-10-0147-01 First Names at Birth by Sex at Birth, Selected Indicators (Number)." <https://doi.org/10.25318/1710014701-eng>.
- University of British Columbia. 2020. "UBC Statement of Financial Information." Open Government - Open Data. <https://finance.ubc.ca/reporting-planning-analysis/reports-and-disclosures>.
- Van Rossum, Guido, and Fred L. Drake. 2009. *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace.