CS 736 Information Visualization

WINTER SEMESTER 2024

INSTRUCTOR: DR. Chris Street

Problem Domain

I am designing to showcase a decade's worth of employment trend data categorized by NOC (National Occupational Classification), which will effectively support professionals working in career development departments within universities or high schools. By providing access to comprehensive historical employment data, the application aims to empower career development officers to offer practical and data-backed advice to students regarding their career paths and educational choices.

Customers: Career Development Advisors in high schools and universities.

Problem to solve: Customers might provide their students with guidance that is supported by rational and practical career options.

Task and Goal: The users look for which industries create the most jobs in each province. So, by displaying the graphic, the user can advise their students on why the career route appears to be successful.

Proposed solution

The proposed solution for a data visualization dashboard shows the top five fastest increasing employment sectors in each province over the last decade. Users can interact with the dashboard by clicking on a provincial card component, which will display time series patterns for those industries. The backend is powered by MongoDB, with data preprocessed with Python and Pandas from Statistics Canada. This approach combines advanced visualization techniques with interactive technologies advocated in the literature

to improve clarity and user engagement.

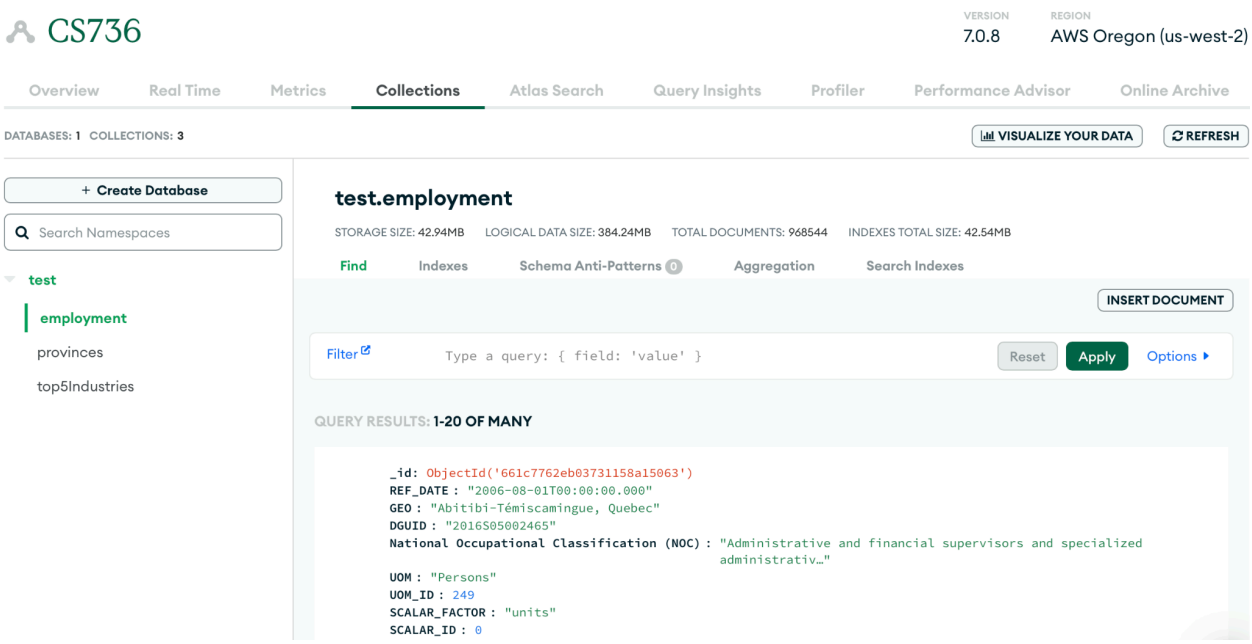
Data Preprocessing

I obtained the data from Statistics Canada for over a 10-year period (2006 January ~ 2024 February), including National Occupation Code (NOC), industrial categories (National Occupation Code, NOC which are predefined by standards), and provincial categories. In order to use the NoSQL database (MongoDB) as my data store and backend, I preprocessed the raw data from Statistics Canada using Python and the Pandas package. The dashboard should display the top five expanding employment industries in each province over the last ten years. Next, the user clicks on a province card, which takes them to the provincial time series (monthly) trends for the top five industries.

Data types:

Employment Value : Ordinal continuous and independent value
National Occupation Code: Nominal Categorical and independent value
Provincial: Nominal Categorical and independent value
Year and Month: Ordinal discrete and independent value

I have three collections on MongoDB Atlas (cloud): one for the dashboard (top5Industries), one for everything (employment), and province for mapping those collections by province.



Data Reference:

Statistics Canada. Statistics Canada. Table 14-10-0431-01 Employment by occupation, economic regions, three-month moving average, unadjusted for seasonality

DOI: https://doi.org/10.25318/1410043101-eng

- **Implement Interactive Tools** like those described in Voyager, allowing users to explore various aspects of the data through automatic visualization recommendations and faceted browsing.

This approach will not only aid in providing a comprehensive understanding of the data but also enhance user engagement and the discovery of insightful patterns. The integration of these methodologies will aim to provide a tool that is both

informative and intuitive, catering to hypothetical users ranging from policymakers to job seekers.

By drawing on high-quality research from established sources in the visualization community, this project will ensure its foundations are scientifically robust, enhancing both its academic and practical value.

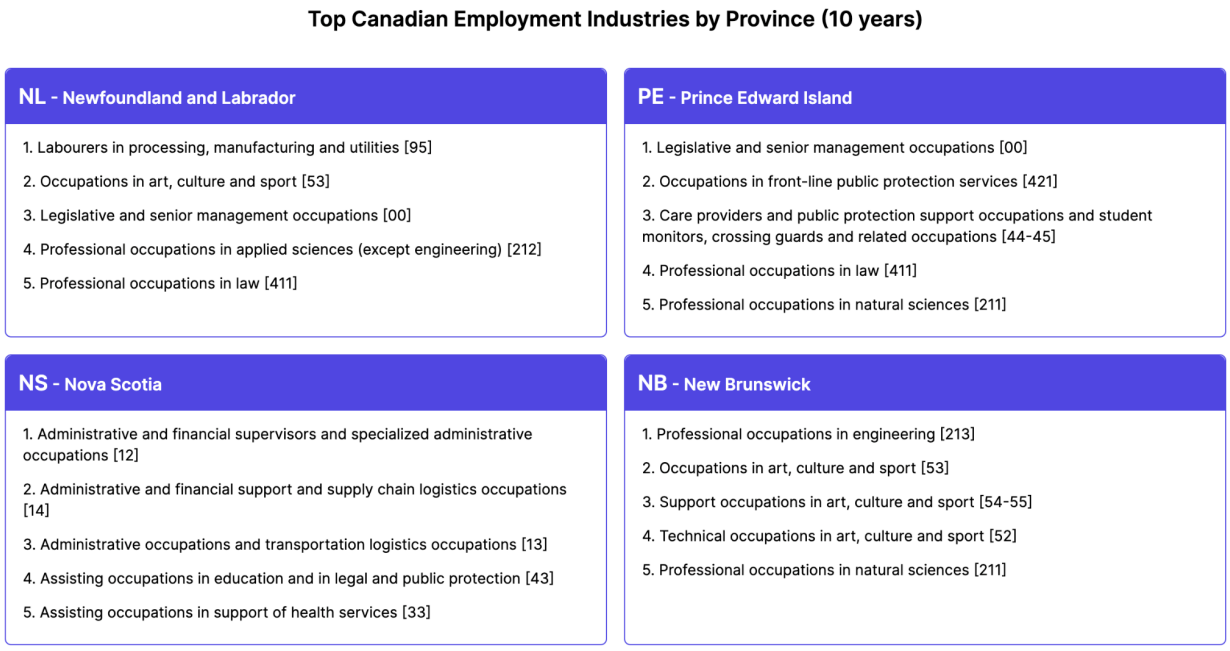


Figure 1. Dashboard

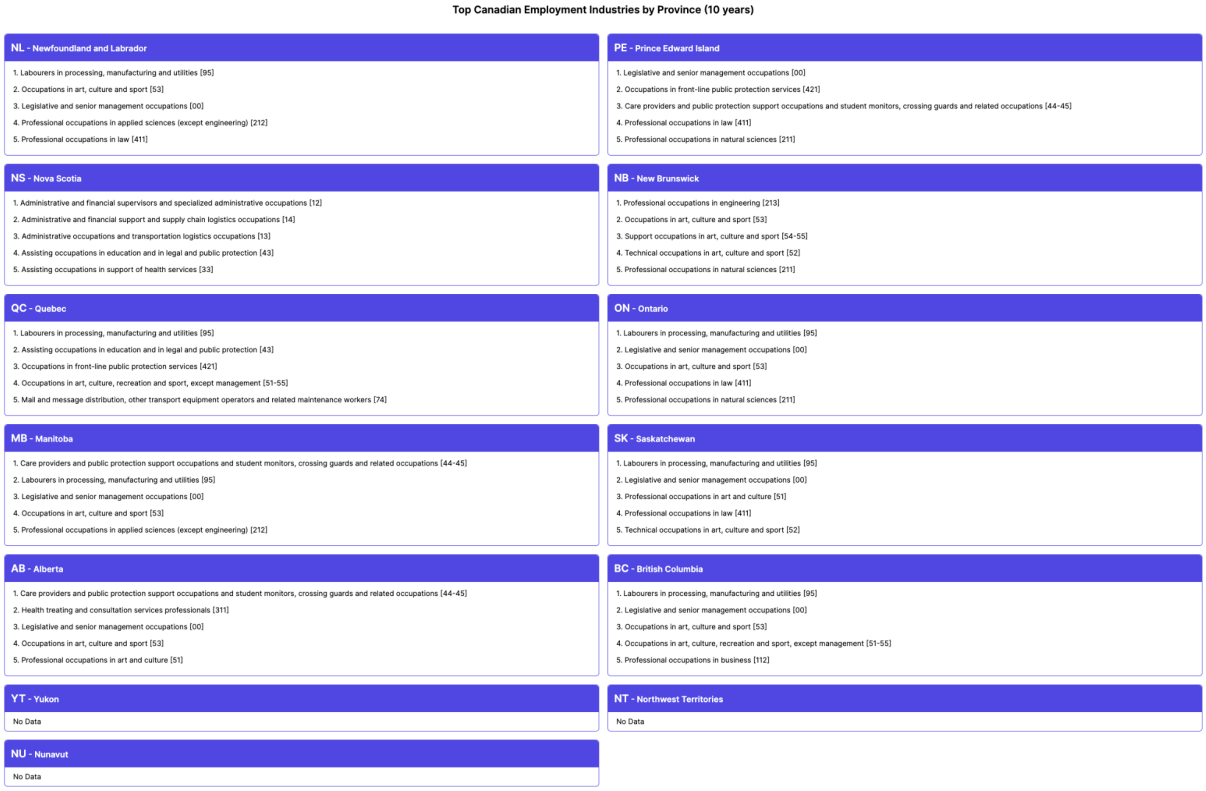


Figure 2. Dashboard as a whole

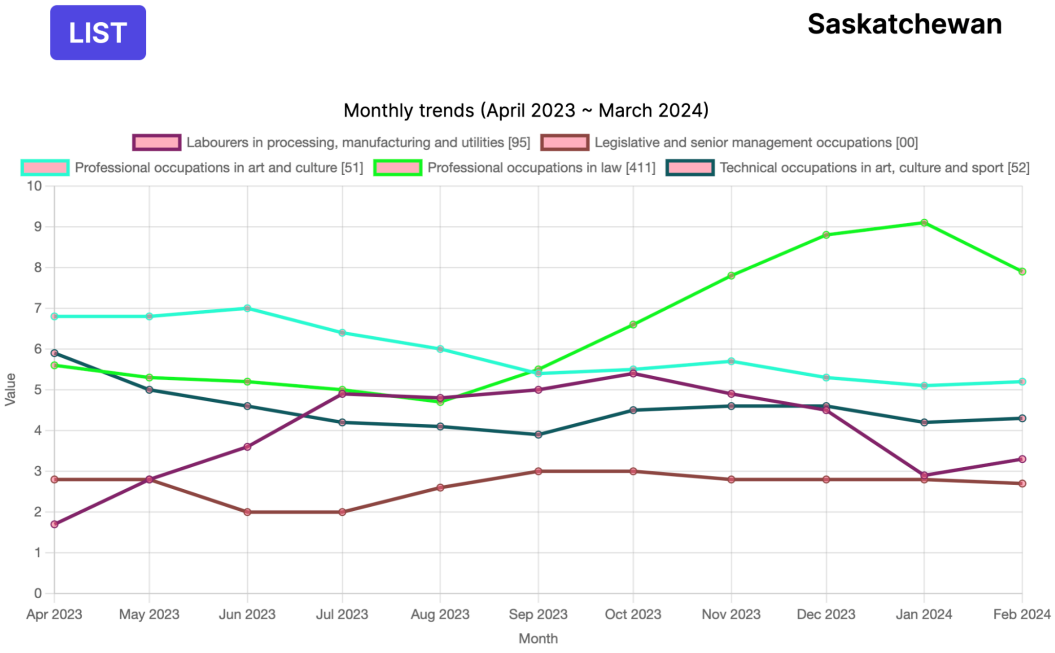


Figure 3. Individual view of the province (Sasktachewan)

- Pre-attentive Processing:

Use of Color: Different colors are used to distinguish between various industries. Color hues are chosen for their pre-attentive qualities, making it easier for users to quickly identify trends across different sectors.

Shape and Size: Various shapes and sizes of markers on the line charts indicate significant data points like peaks, troughs, and anomalies, enabling quick recognition and analysis without in-depth scrutiny.

- Visual Variables:

Position: Time is represented horizontally (x-axis), and employment metrics (like number of jobs) are displayed vertically (y-axis), adhering to Bertin's principle that position is the best variable to represent quantitative data.

- Gestalt Principles:

Proximity and Similarity: Data related to the same industry are grouped closely and use similar colors and shapes, enabling users to perceive them as part of the same group.

Continuity: Lines are used in time series data to indicate continuity, guiding the user's eye across the data points in chronological order.

High Data-to-Ink Ratio:

Minimalism: The design avoids unnecessary decorative elements, focusing on essential data representation only. This principle ensures that the user's attention is drawn directly to the data without distractions.

Scenario of Use

Consider a university career adviser using the dashboard to help a student interested in relocating to British Columbia. The advisor pulls up the province's trend data, which shows that the IT sector has been continually rising. Using this information, the adviser can offer courses and paths that are consistent with the industry trend, allowing the student to make an informed decision about their future.

Discussion on Benefits and Limitations

Benefits

- Provides a reliable, data-driven basis for career advice.
- Enhances the advisor’s ability to track and explain industry trends over time.
- Interactive elements improve user engagement and ease of use.

Limitations

- Deeply coupled with the availability and accuracy of data from Statistics Canada.
- May not capture real-time job market fluctuations due to the data being updated on a monthly basis.