Data Analytics careers: How Much Trouble Are We In?

UofT SCS Data Analytics

Project 2 – Group 6

2019 January 15

# Project Proposal

## Team 6 in the Six Members

Jesse

Sara

Laurel

Jose Tomines

## Project Scope

Analyze the distribution of Data Analytics positions (Data Analyst, Data Scientist, Data Engineer) in Canada by scraping Indeed. We will scrap Glassdoor to get salaries for these positions by position. We will also explore whether these positions are in cities/provinces that are appealing (cost of living) using the Numbeo API.

## Breakdown of Questions

1. How many positions are there per city/province
2. What is the comparison between salary and cost of living

## Potential Data Sources

1. Indeed
2. Glassdoor
3. Numbeo API
4. Stats Canada API

## Potential Data Tables

|  |  |
| --- | --- |
| Job Classification ID (pk) | Job Title |
| 1 | Data Analyst |
| 2 | Data Engineer |
| 3 | Data Scientist |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Posting ID (pk) | Job Classification ID (fk) | Title | City ID (fk) | Province ID | Source ID (fk) |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Province ID (pk) | Province | Cost of Living | Population | Source ID (fk) |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| City ID (pk) | City | Province ID (fk) | Population | Cost of Living | Average Housing Cost | Source ID (fk) |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Salary ID (pk) | Job Classification ID (fk) | GlassDoor Salary | Province ID (fk) | Source ID (fk) |

|  |  |
| --- | --- |
| Source ID (pk) | Source |

fk denotes foreign key

pk denotes primary key

## Breakdown of Tasks

Laurel:

* Scrape Data Analyst, Data Engineer, Data Scientist salaries for 6 provinces to create the Salary Table
* SQL queries of final SQL database to extract information for analytics and plotting.

Sara:

* Scrape Data Analyst, Data Engineer, Data Scientist salaries for 6 provinces to create the Salary Table.
* Cleaning all raw extracted data to get 6 final tables ready for loading onto SQL
* Load to SQL

Jesse:

* Scrape Indeed for Data Analyst, Data Engineer, Data Scientist positions for Canada and extract Title, Company, Salary (if applicable), City, Province
* Load to SQL

Jose:

* Use stats Canada and Numbeo API to find information such as population, cost of living, etc. to create the City and Province Table
* Final Report to summarize all finding

# Approval and Authority to Proceed

We approve the project as described above, and authorize the team to proceed.

|  |  |  |
| --- | --- | --- |
| Name | Title | Date |
|  |  |  |
|  |  |  |
|  |  |  |