Midterm Status Presentation for The Analysis of Rural Business Performance

Tingwen Hua, Bowen Yang,
Yilin Sun, Song(Alice) Zhang

May 31st, 2022

Overview

- Team Introduction
- Client Introduction
- Datasets
- Research Questions
- Objective
- Achievements
- Technologies
- Roadblock

Team Introduction

Tingwen Hua

Academic background:Pepperdine University, major in Business and Math

Working background: Business Analytics in Shanghai for 2 years

Hobby: Lego, Travel

Team Introduction

Bowen Yang

Academic background: Simon Fraser University, major in Computing Science

Working background: Software Development Engineer, web application development, cyber security

Live in Vancouver

Team Introduction

Yilin Sun

Academic Background:

University of Pennsylvania, major in Computer Science

Georgetown University, major in Asian Studies

Working background:

SDE Intern at Intuit

Team Introduction

Song (Alice) Zhang

Academic Background:

Jilin University, PhD in Mathematical Statistics

Working Background:

Postdoc at Dalian Commodity Exchange, research on options portfolio risk management

Research Analyst at Shanghai Clearing House, research on CCP clearing and bond market

Client Introduction

Organization: Statistics Canada

Statistics Canada is the national statistical office. The agency ensures Canadians have the key information on Canada's economy, society and environment that they require to function effectively as citizens and decision makers.

Managers: Harris Jafri, Eric Baxter and Mahdia Khodja

Topic: Rural Canada Business

Datasets

Rural Canada Business Profile from 2017 to 2019

- Main focus: Small Business and Medium Business Financial Information
- Variables:
    - Location: Rural/Urban
    - Region: Provinces
    - Incorporation Status: Incorporated/Unincorporated
    - Year: 2017 to 2019
    - Industries: North American Industry Classification System, NAICS

Research Questions

1. **How does the Canadian rural business perform according to different dimensions/variables?**

2. **How can the RCBP data be connected with other StatCan data to produce valuable analysis?**

Objective

In this project, we intend to achieve 3 main objectives:

1. Exploratory analysis of the Rural Canada Business Profiles (RCBP) database.

2. Interactive dashboards:

    Rural Canada Business Profiles (RCBP) dashboard

    Overview dashboard

3. Cross-analysis of RCBP with other public Statistics Canada data (open ended) .

Timeline

Week	Theme	Goals
1 (01 – 09 May)	Proposal	Establish regular connection with client, understand the problem, develop the work plan, deliver project proposal and presentation
2 (10 – 17 May)	EDA and data visualization	Understand the RCBP dataset, data wrangling, do EDA and data visualization, develop analytical aspects
3 (18 – 24 May)	Data visualization and Analytical paper (first draft)	Data visualization, deliver the first draft of the analytical paper
4 (25 – 31 May)	Dashboard Mid-project presentation	Develop the dashboard, make the mid-project presentation
5 (01 – 07 June)	Dashboard	Develop the dashboard, debugging
6 (08 – 14 June)	Open-ended Question	Explore other related public StatCan data, do the cross-analysis with RCBP
7 (15 – 21 June)	Final report Final Presentation	Deliver the Analytical paper, final report and make the final presentation

Achievements

Analytical paper — first draft

- [Analysis of Rural Business Performance](#)

RCBP Dashboard

- [Canada Medium Business Profile from 2017 to 2019](#)
- [Canada Small Business Profile from 2017 to 2019](#)
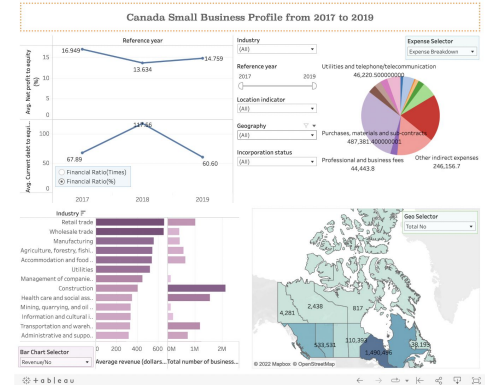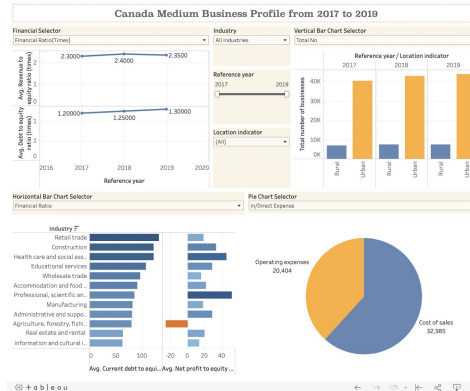
Technologies

Tableau/Altair:

- RCBP Dashboard
- Exploratory Data Analysis
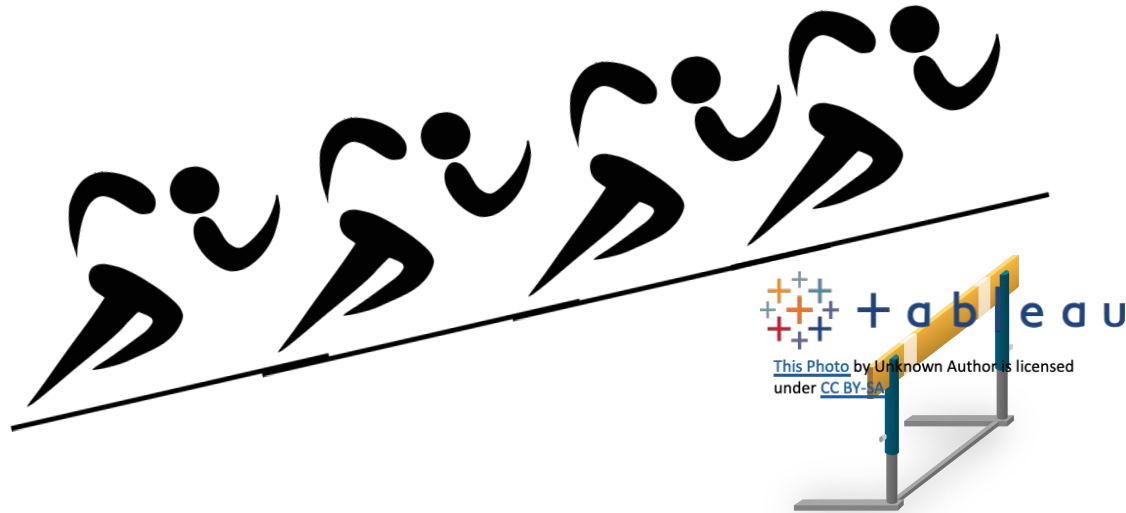- Analytical Paper

Python Pandas:

- Data Loading
- Data Cleaning
- Data Wrangling

```python
def tab_reader(filepathpattern,tabnum,skiprow,low_memory=False):
    tabs = []
    for i in range (tabnum):
        cur_index = i+1
        cur_filepath = filepathpattern.format(cur_index)
        try:
            tab = pd.read_csv(cur_filepath, skiprows=skiprow, low_memory=False)
        except:
            tab = pd.read_csv(cur_filepath, skiprows=skiprow,encoding='latin-1', low_memory=low_memory)
        tabs.append(tab)
    return tabs
```

```python
# 2017 data
rcbp_2017_mb_profitMargin_Tabs = tab_reader("2017_csv_eng/2017_Medium businesses_Profit margin based_csv/_2017_Medium businesses_Profit margin_Tab{}.csv",5,5)
rcbp_2017_mb_revenue_Tabs = tab_reader("2017_csv_eng/2017_Medium businesses_Total revenue based_csv/_2017_Eng_Medium_Revenue_Tab{}.csv",5,5)
rcbp_2017_sb_profitMargin_Tabs = tab_reader("2017_csv_eng/2017_Small businesses_Profit margin based_csv/_2017_Small businesses_Profit margin based_Tab{}.csv",7,5)
rcbp_2017_sb_profitMargin_Tabs[0] = tab_reader("2017_csv_eng/2017_Small businesses_Profit margin based_csv/_2017_Small businesses_Profit margin based_Tab{}.csv",1,6)[0]
rcbp_2017_sb_revenue_Tabs = tab_reader("2017_csv_eng/2017_Small businesses_Total revenue based_csv/_2017_Small businesses_Total revenue based_Tab{}.csv",7,5)
rcbp_2017_sb_revenue_Tabs[0] = tab_reader("2017_csv_eng/2017_Small businesses_Total revenue based_csv/_2017_Small businesses_Total revenue based_Tab{}.csv",1,6)[0]
```

Roadblock

Use Tableau to create RCBP dashboard

Thank You!