



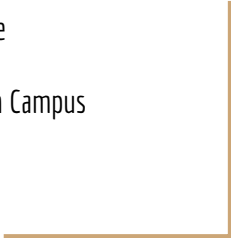
# A Proposal for The Analysis of Rural Business Performance

Tingwen Hua, Bowen Yang, Yilin Sun, Song Zhang

Master of Data Science Capstone

University of British Columbia Okanagan Campus

2022/05/09



# Agenda

- Self-Introduction
- Client & Project Introduction
- Objectives
- Timeline & Workflow
- Tasks Completed & Next week plan

# Self 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

# Self Introduction

Tingwen Hua

Academic background: Pepperdine University,  
major in Business and Math

Working background: Business Analytics in  
Shanghai for 2 years

Hobby: Lego, Travel



# Self Introduction

Yilin Sun

Academic Background:

University of Pennsylvania, major in Computer Science

Georgetown University, major in Asian Studies

Working background:

SDE Intern at Intuit

# Self Introduction

Song 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

Managers: Harris Jafri, Eric Baxter and Mahdia Khodja

Topic: Rural Canada Business

Dataset: Rural Canada Business Profiles 2017-2019

# Project Introduction: Dataset

## Rural Canada Business Profiles

Counts and financial data on small businesses ( $\$30,000 < \text{yearly revenue} < \$5 \text{ million}$ ) and medium businesses ( $\$5 \text{ million} < \text{yearly revenue} < \$20 \text{ million}$ ) in Canada, aggregated by geography, industry, incorporation status, and some other relevant variables.

Year: 2017-2019



# Project Introduction: Dataset

For each business size, it provides the following financial report as tables:

- Counts of businesses
- Revenue breakdowns
- Expense breakdowns
- Balance sheet items (assets, liabilities, equity breakdowns)
- Financial ratios

In each table, the data are managed by the following variables:

- Small or medium size of business by annual revenues
- Various levels of geography (Canada, region, province/territory)
- Rural and urban areas
- Industry
- Incorporation status
- Profitable and non-profitable businesses

# Project Introduction: Previous Related Work

Our Client published a research paper in March focusing on revenue performance of rural area in contrast to urban using the RCBP dataset.

Rural and Small Town Canada Analysis Bulletin

## A profile of businesses in rural Canada, 2017 to 2019

by [Mahdia Khodja](#), [Eric Baxter](#) and [Haaris Jafri](#)

Release date: March 11, 2022



[More information](#)

[PDF version](#)

# Objectives

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

1. To provide an exploratory analysis of the Rural Canada Business Profiles (RCBP) database.
2. To implement an interactive dashboard that allows users to choose dimensions/variables and get related visualizations and detailed links.
3. To perform a cross-analysis of RCBP with other public Statistics Canada data (open ended)

# 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?

# Methods - First Objective

- Loading data to pandas dataframe
- Data cleaning
- Data wrangling
- Data exploration
- Data visualization
- Summarizing results of exploration to analysis report

# Methods - Second Objective

- Understanding datasets used on Rural Canada Statistics Portal page
- Comparing Dash in Python and Tableau
- Generating easy-to-understand visualization
- Implement Interactive Dashboard

# Methods - Third Objective

- Performing cross-analysis on RCBP and other datasets outside the rural hub.
- Investigating the relationship between variables in these datasets through linear/multi-linear regression.
- Applying predictors selection methods (like Forward Selection/Backward Selection, Ridge Regression and LASSO) for selecting “more significant variables”.

# Deliverables

- Analytical paper

An exploratory analysis report of the Rural Canada Business Profiles (RCBP) database with supporting visualizations.

- Dashboard

An interactive dashboard for the Rural Canada Statistics Portal that could be used as an overview page.

- Final report

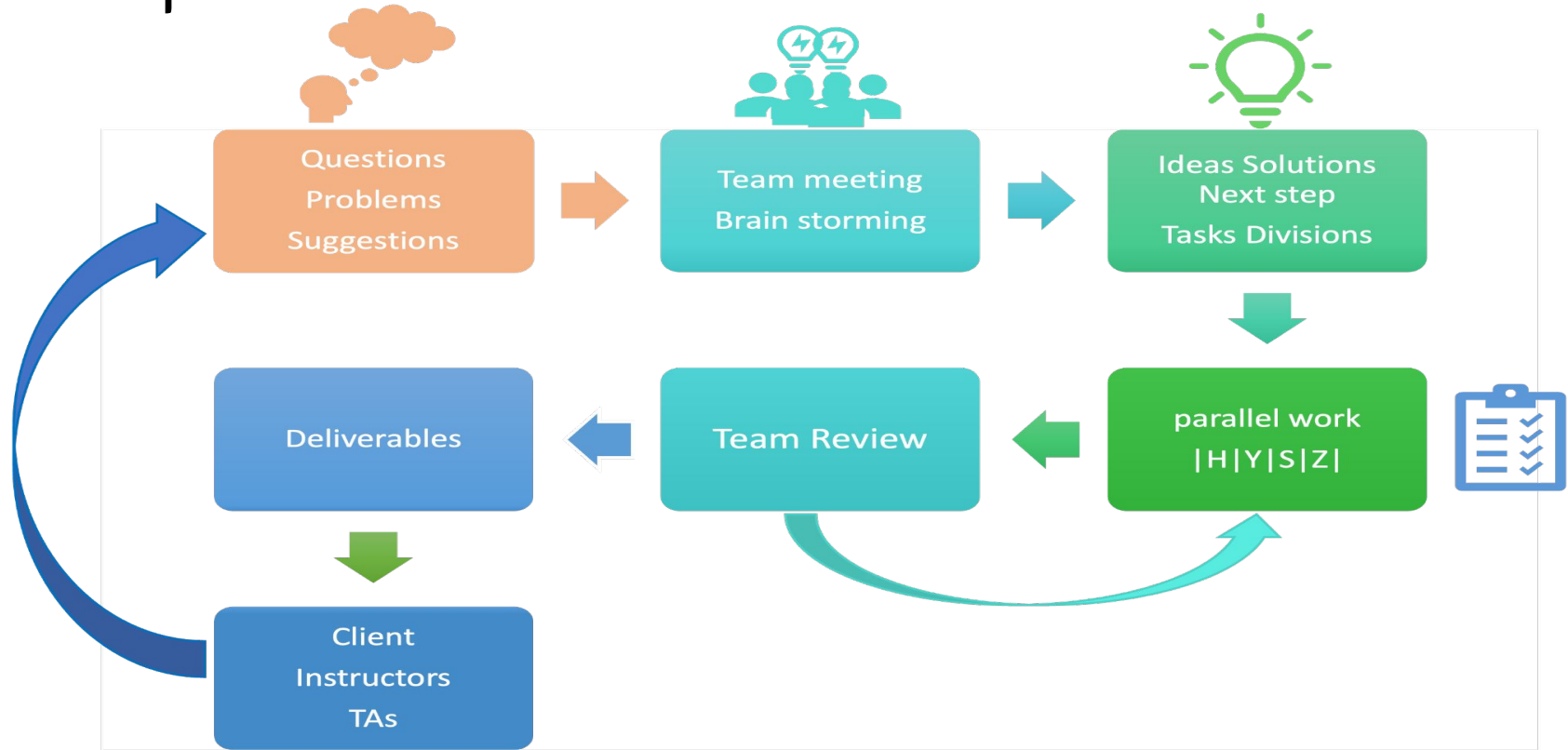
A report that contains all our works, which includes the analysis of RCBP database, the display of the dashboard, and the insight analysis of the open-ended questions about cross-analysis of RCBP.



# 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
<b>4 (25 – 31 May)</b>	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
<b>7 (15 – 21 June)</b>	Final report Final Presentation	Deliver the Analytical paper, final report and make the final presentation

# Workflow

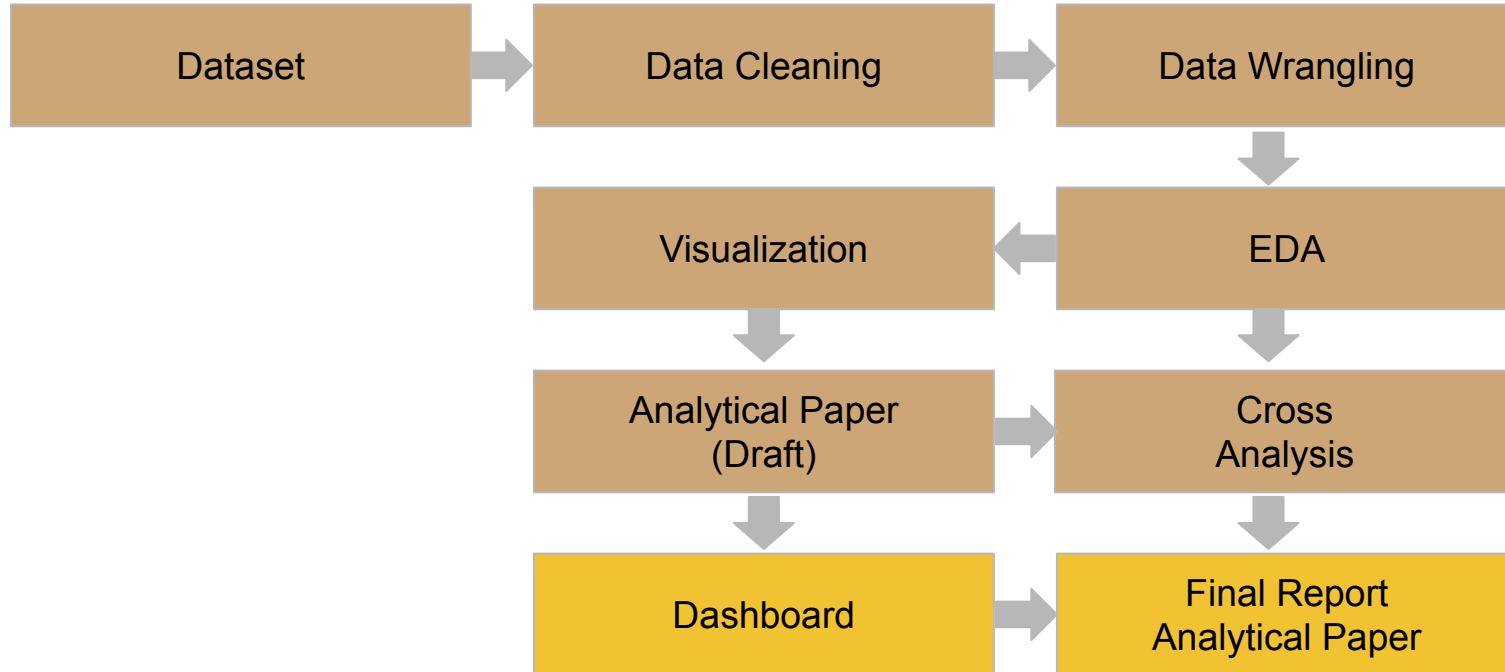


# Collaboration

Our team will work collaboratively and make the most of everyone's strengths:

- Meeting minutes — in turn
- Tingwen Hua — Client liaison
- Tingwen Hua — more business ideas
- Bowen Yang — more coding work
- Yilin Sun — more coding work
- Song Zhang — more modeling work

# Processing



# Tasks completed

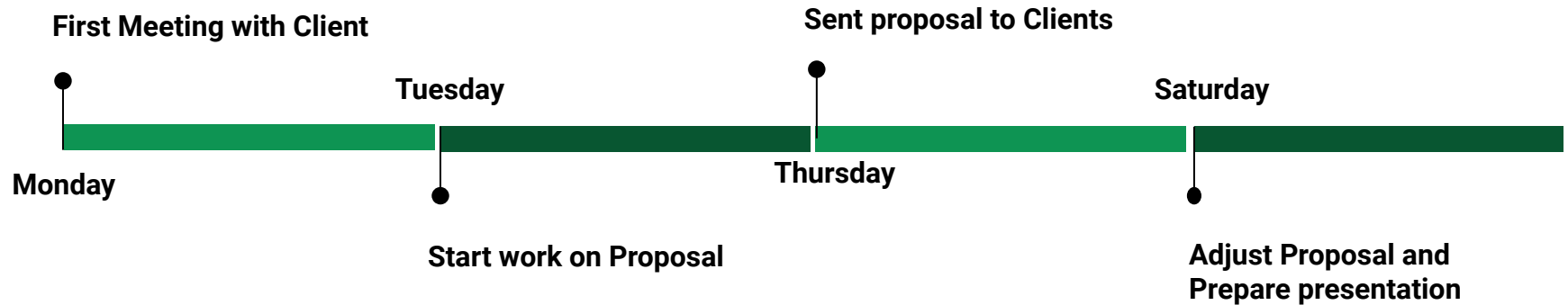
## Communications:

1. First Meeting with Client: Monday May 2nd 1 hr
2. Several Meetings within team to discuss work distribution and progress:
  - Tuesday May 3rd: Proposal discuss 2 hr
  - Wednesday May 4th: Review first draft of proposal 3hr
  - Thursday May 5th: Proposal completion & Presentation Distribution 2hr
  - Saturday May 7th: Adjust proposal based on client's feedback & review slides 1.5hr
  - Monday May 9th: Presentation Mock 1hr
3. Get Feedback from Clients on the proposal: Friday May 6th through email

# Tasks completed

## Products:

- Proposal of team project
- First week report presentation



# Next Week:

1. Weekly meeting with Clients
2. Data Processing
  - a. Data cleaning
  - b. Data wrangling
  - c. Data visualization
3. EDA and develop analytical aspects



Thank You!

