Kingdom of Saudi Arabia

المملكة العربية السعودية

جامعة
الملك سعود
King Saud University
1957

Exploring the Depths of Data Science: A Comprehensive Journey with Rust, Docker, and Apache Arrow

Salman Almuzaini

Data Odyssey: A Personal Journey through Analysis and Code

King Saud University

November, 2023

Table of Contents

Summary

This document is a detailed record of my foray into data science with Rust, capturing the essence of the journey from initial setup using Cargo to tackling various challenges. It narrates the progression through stages that involved integration with APIs, meticulous error management, and navigating Docker for environmental stability, emphasizing the intricacies of cross-compiling and static linking.

Throughout this exploration, data played a central role, with CSV files such as "Table 2.4.4U. Price Indexes for Personal Consumption Expenditures by Type of Product" and "Table 2.6. Personal Income and Its Disposition, Monthly" serving as fundamental resources. These files, derived from the Bureau of Economic Analysis (BEA), were crucial in analyzing economic indicators within the Rust environment.

A notable point in the journey was the division of the "Personal Consumption Expenditures Price Index.xlsx" file into individual CSV sheets such as "Contents.csv", "Table 1.csv", through "Table 7.csv", a process facilitated by a Python script sourced from GitHub. This script, `getsheets.py`, was instrumental in transforming the multifaceted Excel document into a structured array of CSV files, thereby simplifying the dataset for subsequent analysis.

The narrative also highlights the strategic shift to a more manageable dataset, 'organization-100.csv', for effective analysis, and the utilization of 'iris.csv' for exploring data processing capabilities within Rust. Additionally, the Rust Jupyter notebook, sourced from the Rust user's forum and adapted from online tutorials, provided an interactive platform for executing and testing code snippets, further enriching the learning experience.

The document reflects on the importance of adaptability in learning and the precision required for robust error handling in Rust. It is a testament to the technical milestones achieved and an introspection of the strategic decision-making and problem-solving approaches vital for data science within the Rust ecosystem.

N.1 Day 1: Rust, APIs, and Containerization Challenges