Essential Softwares for Data Science and Analytics Foundation Course

Dr. Austin Brown

Kennesaw State University

Why this Course?

- **Demand for Data Skills**: The demand for data science and analytics skills is growing across all industries.
- **Versatility of Tools**: Learning multiple tools like SAS, R, and Python provides flexibility and increases employability.
- **Comprehensive Skill Set**: This course covers essential foundational aspects such as data importation, transformation, analysis, and visualization.
- **Foundation for Advanced Learning**: Provides a strong foundation for more advanced data science and analytics topics.

Course Objectives

- **Familiarize with Key Software**: Navigate and use SAS Studio, RStudio, and Python IDEs.
- **Data Handling**: Import, clean, and preprocess data from various sources.
- **Basic Analysis and Visualization**: Perform simple data transformations, summaries, and visualizations.
- **Output and Reporting**: Export results in various formats and create reproducible reports.
- **Practical Application**: Apply skills in a comprehensive project, simulating real-world data scenarios.

Course Outline

1. **Introduction to Software Environments**
2. **Data Importation**
3. **Data Entry**
4. **Understanding Column Contents**
5. **Simple Data Transformations/Queries**
6. **Simple Descriptive Methods**
7. **Data Preprocessing**
8. **Outputting Results**

Introduction to Software Environments

- **SAS Studio**
  - Overview and installation
  - Basic navigation and features
- **RStudio**
  - Overview and installation
  - Basic navigation and features
- **Python IDEs**
  - Google Colab, Jupyter Notebook, PyCharm, and other popular IDEs
  - Overview and installation
  - Basic navigation and features

Introduction to SAS Studio

- **Overview and Installation**: SAS Studio is a web-based application that provides access to SAS functionality. It can be installed locally on your machine or it can be accessed through the web via SAS OnDemand for Academics.
- **Basic Navigation**: SAS Studio interface includes features such as code editor, log window, and task pane for accessing SAS tasks.
- **Key Features**: Features include data import/export, programming with SAS code, interactive data exploration, and visual analytics.

Introduction to RStudio

- **Overview and Installation**: RStudio is an integrated development environment (IDE) for R programming. It can be installed as a standalone application or accessed via Posit Cloud.
- **Basic Navigation**: RStudio interface includes panes for scripts, console, environment, and plots. It also provides access to R packages and help documentation.
- **Key Features**: Features include code editor with syntax highlighting, integrated R console, package management, and version control integration.

Introduction to Python IDEs (Using Google Colab)

- **Overview and Installation**: Python offers various IDEs such as Jupyter Notebook, PyCharm, and Spyder. Each IDE has its installation process. Google Colab is a cloud-based Jupyter notebook environment that allows you to write and execute Python code in the browser. It provides access to Google's powerful hardware for free and requires no installation.
- **Basic Navigation**: Google Colab provides a familiar Jupyter notebook interface with cells for code, text, and outputs. It supports collaborative editing and version control.
- **Key Features**: Features include integration with Google Drive for storage, support for Python libraries and packages, GPU/TPU acceleration, and easy sharing of notebooks.Also, include code autocompletion, inline documentation, variable explorer, and integration with data science libraries like pandas, numpy, and scikit-learn.

Hands-On Practice

- **Exercise**: Access SAS Studio, RStudio, and Google Colab.
- **Explore**: Let familiarize ourselves with the interface of each environment.
- **Try**: Writing a simple code snippet in each environment to perform basic operations like printing “Welcome to KSU!” or calculating simple arithmetic expressions.