New Era University
College of Informatics and Computing Studies
Department of Computer Science

FREE ELECTIVE 3

CCSEL3-18

Data Science Tools

by: **Cognitive Class**

ONLINE COURSE DOCUMENTATION

Marasigan, Vem Aiensi A.
BS Computer Science | 21-11295-310

Module 1: Language of Data Science

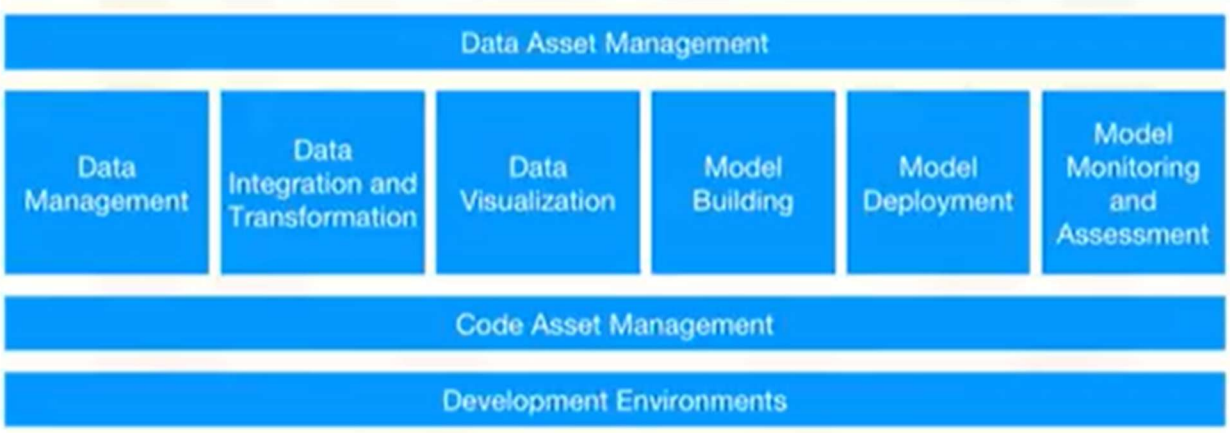
[Completed by October 6]

Basically, this module introduces languages used in Data Science. These are Python, R, and SQL. Not much is new here except for R since it wasn't tackled before. Apart from that is the difference between Free and Open Source software, which is vaguely discussed in my opinion. Lastly mentioned are the other Languages, although not as popular as the aforementioned, but still exists in the ecosystem of Data Science.

Module 2: Data Science Tools

[Completed by October 22]

This covered quite a lot of tools for data science. Due to the nature of data science, involving many processes, from tranformation to visualization, etc. there are many tools that already exists to aid data scientist. These can be group in categories which are:



There are quite a lot of tools and among those tools mentioned, the ones I personally have interests or had experience are PostgreSQL, Jupiter Notebooks, and Tensorflow for Machine Learning. Other open source tools not mentioned though are Tableu Public or even the simplest tools such as Google Spreadsheet.

Just some examples from the module

Aside from open source tools, there are commercial ones too offering features that can cater all data science workflows. Some Highlighted here is:

This commercial tool is also a cloud based platform.

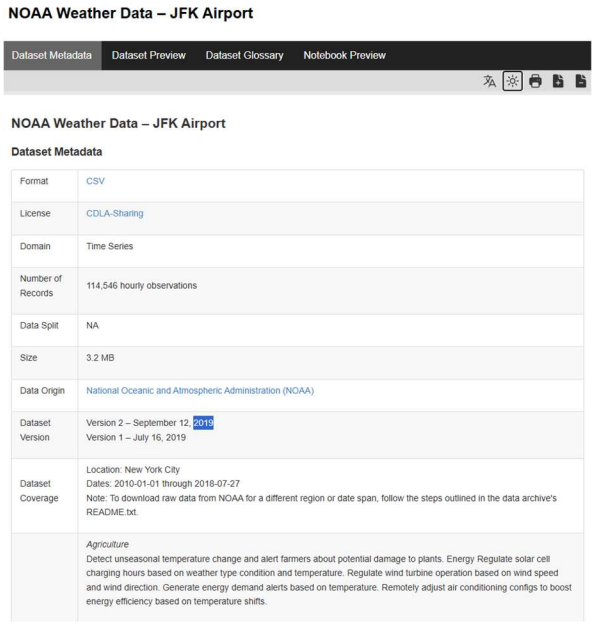
Watson Studio, together with Watson OpenScale, covers the complete development life cycle for all data science, machine learning, and AI tasks

Module 3: Packages, APIs, Data Sets and Models

[Completed by October 22]

This module covered libraries used in data science along with the tools necessary to use those libraries. Another thing is the REST APIs which can be used to access pretrained or existing models, offering convinience to data scientist. Another thing to consider here are the licenses involved in using in using open source data sets and models. Here is some

This module also offered a simple laboratory activity to try out data sets and jupyter notebooks



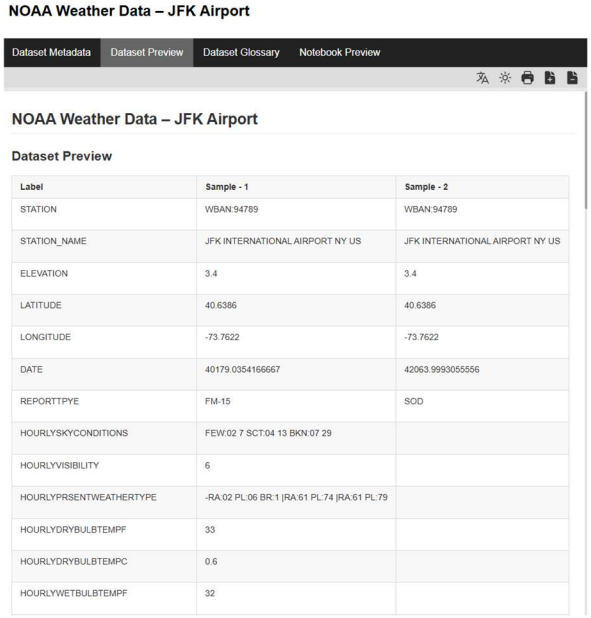
NOAA Weather Data – JFK Airport

Dataset Metadata | Dataset Preview | Dataset Glossary | Notebook Preview

NOAA Weather Data – JFK Airport

Dataset Metadata

Format	CSV
License	CDLA-Sharing
Domain	Time Series
Number of Records	114,546 hourly observations
Data Split	NA
Size	3.2 MB
Data Origin	National Oceanic and Atmospheric Administration (NOAA)
Dataset Version	Version 2 – September 12, 2019 Version 1 – July 16, 2019
Dataset Coverage	Location: New York City Dates: 2010-01-01 through 2018-07-27 Note: To download raw data from NOAA for a different region or date span, follow the steps outlined in the data archive's README.txt.
	Agriculture Detect unseasonal temperature change and alert farmers about potential damage to plants. Energy Regulate solar cell charging hours based on weather type condition and temperature. Regulate wind turbine operation based on wind speed and wind direction. Generate energy demand alerts based on temperature. Remotely adjust air conditioning configs to boost energy efficiency based on temperature shifts.



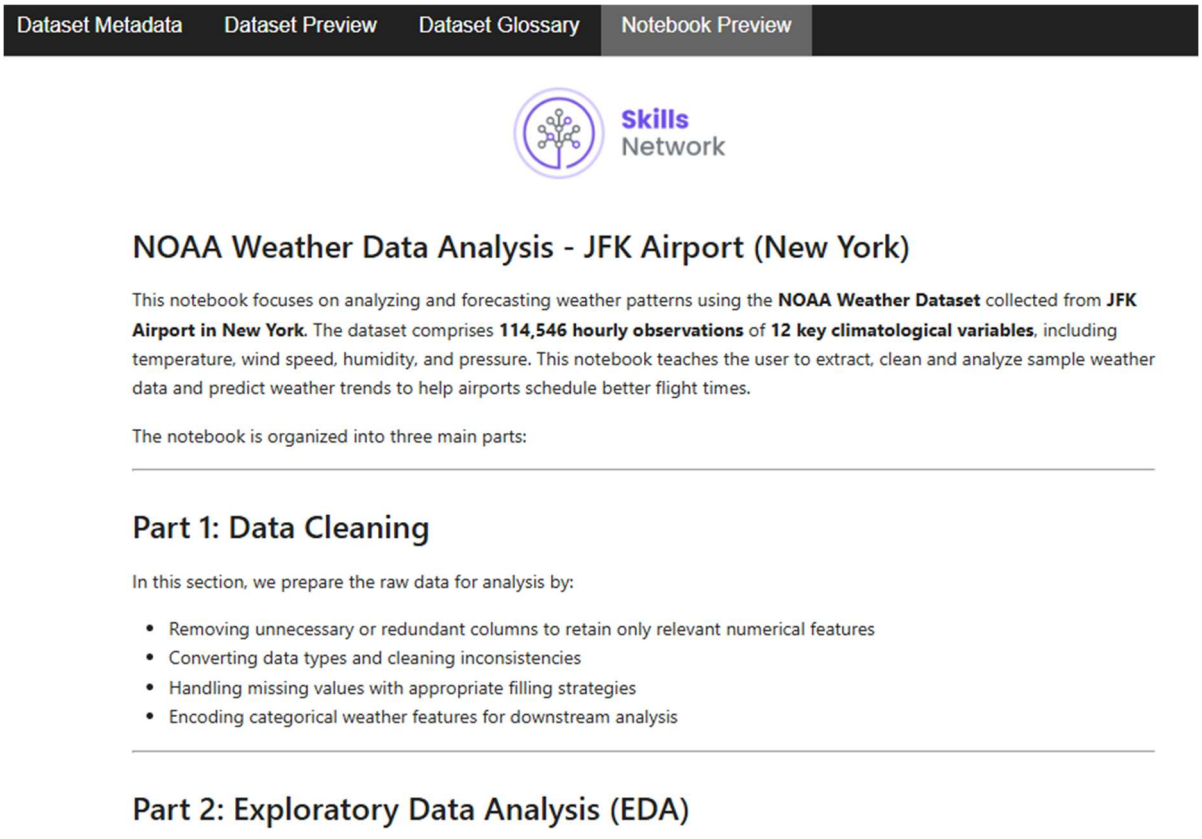
NOAA Weather Data – JFK Airport

Dataset Metadata | Dataset Preview | Dataset Glossary | Notebook Preview

NOAA Weather Data – JFK Airport

Dataset Preview

Label	Sample - 1	Sample - 2
STATION	WBAN:94789	WBAN:94789
STATION_NAME	JFK INTERNATIONAL AIRPORT NY US	JFK INTERNATIONAL AIRPORT NY US
ELEVATION	3.4	3.4
LATITUDE	40.6386	40.6386
LONGITUDE	-73.7622	-73.7622
DATE	40179.0354166667	42063.9993055556
REPORTTPYE	FM-15	SOD
HOURLYSKYCONDITIONS	FEW:02 7 SCT:04 13 BKN:07 29	
HOURLYVISIBILITY	6	
HOURLYPRSENTWEATHERTYPE	-RA:02 PL:06 BR:1 |RA:61 PL:74 |RA:61 PL:79	
HOURLYDRYBULBTEMPF	33	
HOURLYDRYBULBTEMPC	0.6	
HOURLYWETBULBTEMPF	32	



Dataset Metadata | Dataset Preview | Dataset Glossary | Notebook Preview

Skills Network

NOAA Weather Data Analysis - JFK Airport (New York)

This notebook focuses on analyzing and forecasting weather patterns using the **NOAA Weather Dataset** collected from **JFK Airport in New York**. The dataset comprises **114,546 hourly observations** of **12 key climatological variables**, including temperature, wind speed, humidity, and pressure. This notebook teaches the user to extract, clean and analyze sample weather data and predict weather trends to help airports schedule better flight times.

The notebook is organized into three main parts:

Part 1: Data Cleaning

In this section, we prepare the raw data for analysis by:

- Removing unnecessary or redundant columns to retain only relevant numerical features
- Converting data types and cleaning inconsistencies
- Handling missing values with appropriate filling strategies
- Encoding categorical weather features for downstream analysis

---

Part 2: Exploratory Data Analysis (EDA)

Module 4 & 5: Git & Jupyter Notebooks

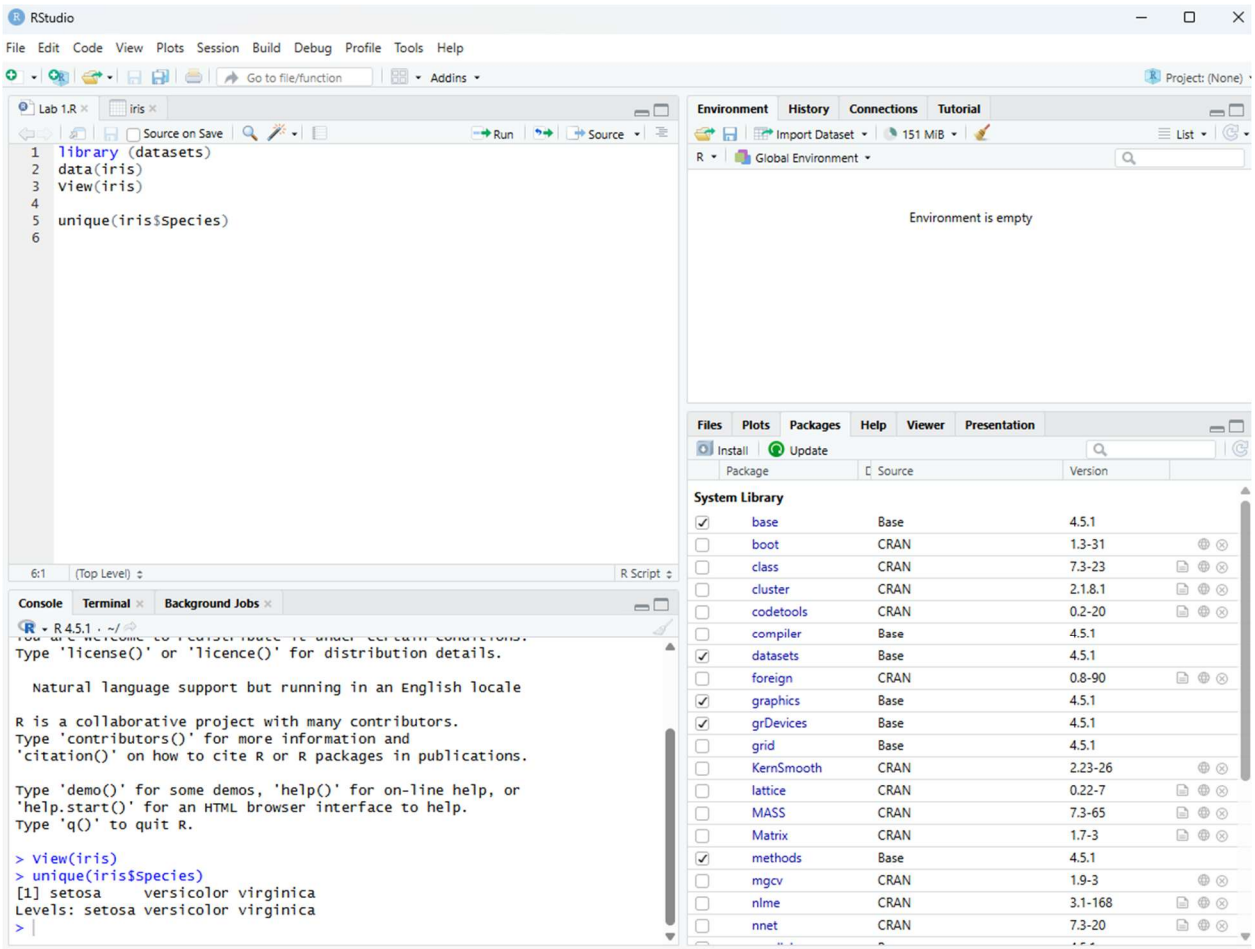
[Completed by October 22]

This module is pretty straightforward. Fundamentals of Git and Jupyter. It also covered GitLab and JupyterLab as well as practice labs.

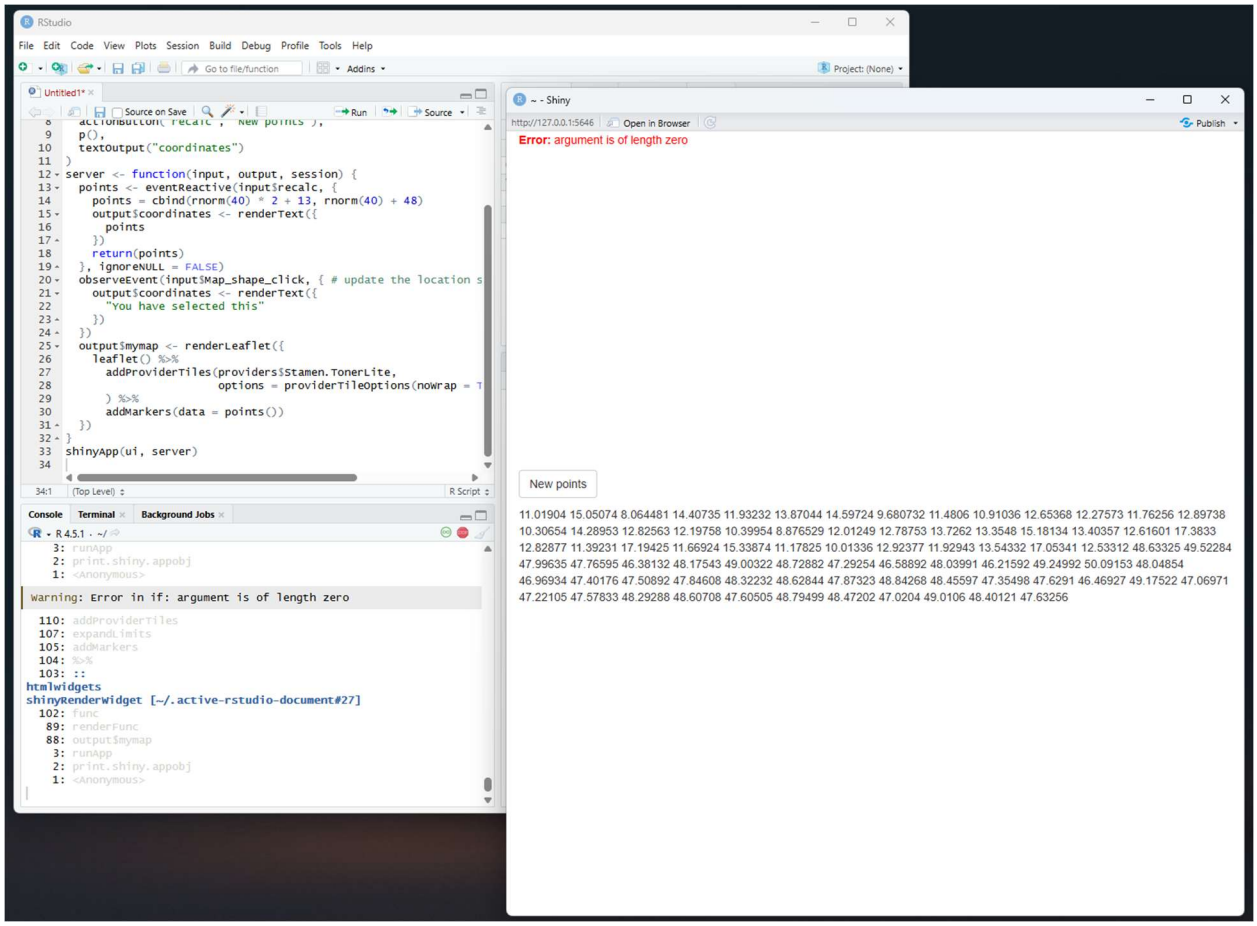
Module 6: R Studio

[Completed by October 22]

The module offered a lab for R Studio so since this would be my first time, I checked it out. This is laboratory 1 for R Studio:



The second lab however doesn't quite work as the module expects it to. I followed the instructions, copied the code exactly as it is but it's kinda off. So, maybe I'll come back to this later. Here is what I get though:



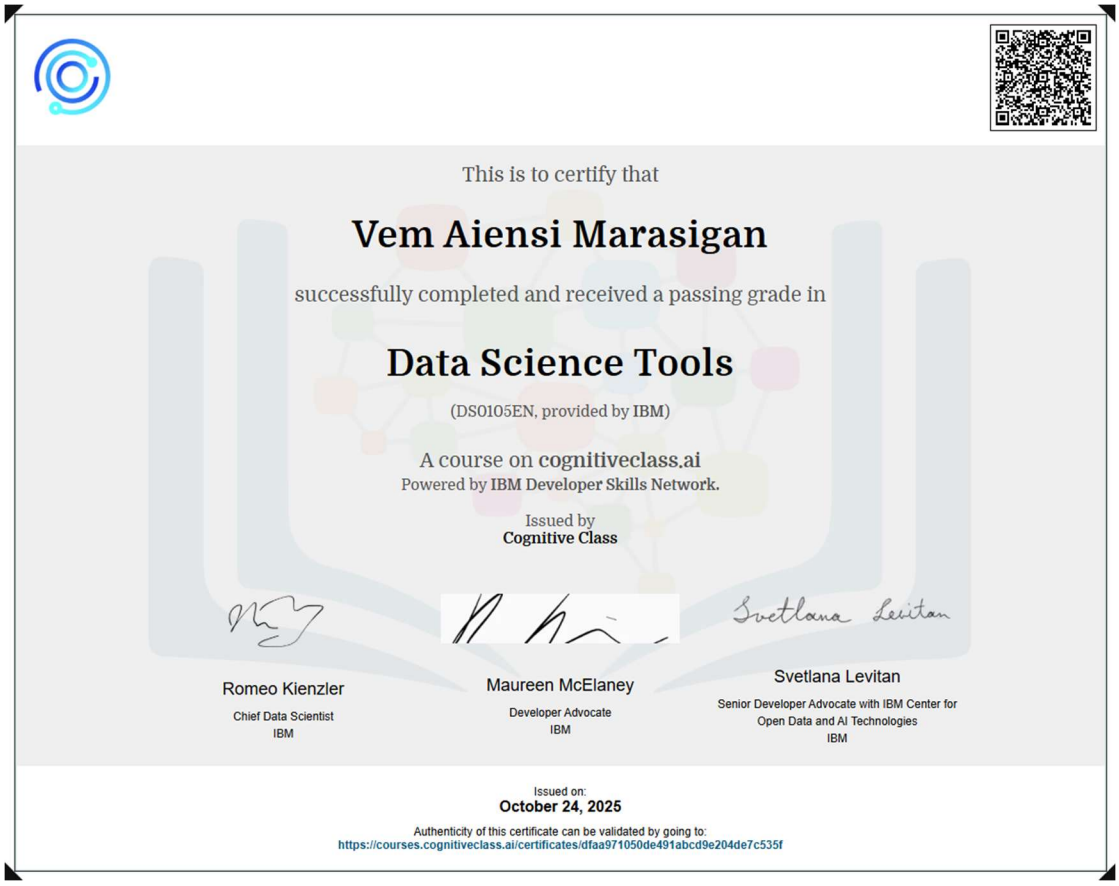
The proceeding instructions are also kinda hard to understand and can't tell if I'm doing it right considering I haven't got the expected outcome in the first place. Maybe it's best to try other tutorials for R besides this one. Still, it's a pretty descent module. It's a beginner course anyway.

Module 6: Watson Studio

[Completed by October 24]

Unfortunately, as much as I love to, I cannot do any of the lab activities as my IBM Free Tier Account has already expired as I've already used it when learning DB2. Although we did managed to pass the Quiz!

COMPLETION RECORD



This is to certify that

Vem Aiensi Marasigan

successfully completed and received a passing grade in

Data Science Tools

(DS0105EN, provided by IBM)

A course on cognitiveclass.ai
Powered by IBM Developer Skills Network.

Issued by
Cognitive Class

Romeo Kienzler
Chief Data Scientist
IBM

Maureen McElaney
Developer Advocate
IBM

Svetlana Levitan
Senior Developer Advocate with IBM Center for Open Data and AI Technologies
IBM

Issued on:
October 24, 2025

Authenticity of this certificate can be validated by going to:
https://courses.cognitiveclass.ai/certificates/dfaa971050de491abcd9e204de7c535f