

Introduction to Data Science with Python

Nicholas Russell Saerang



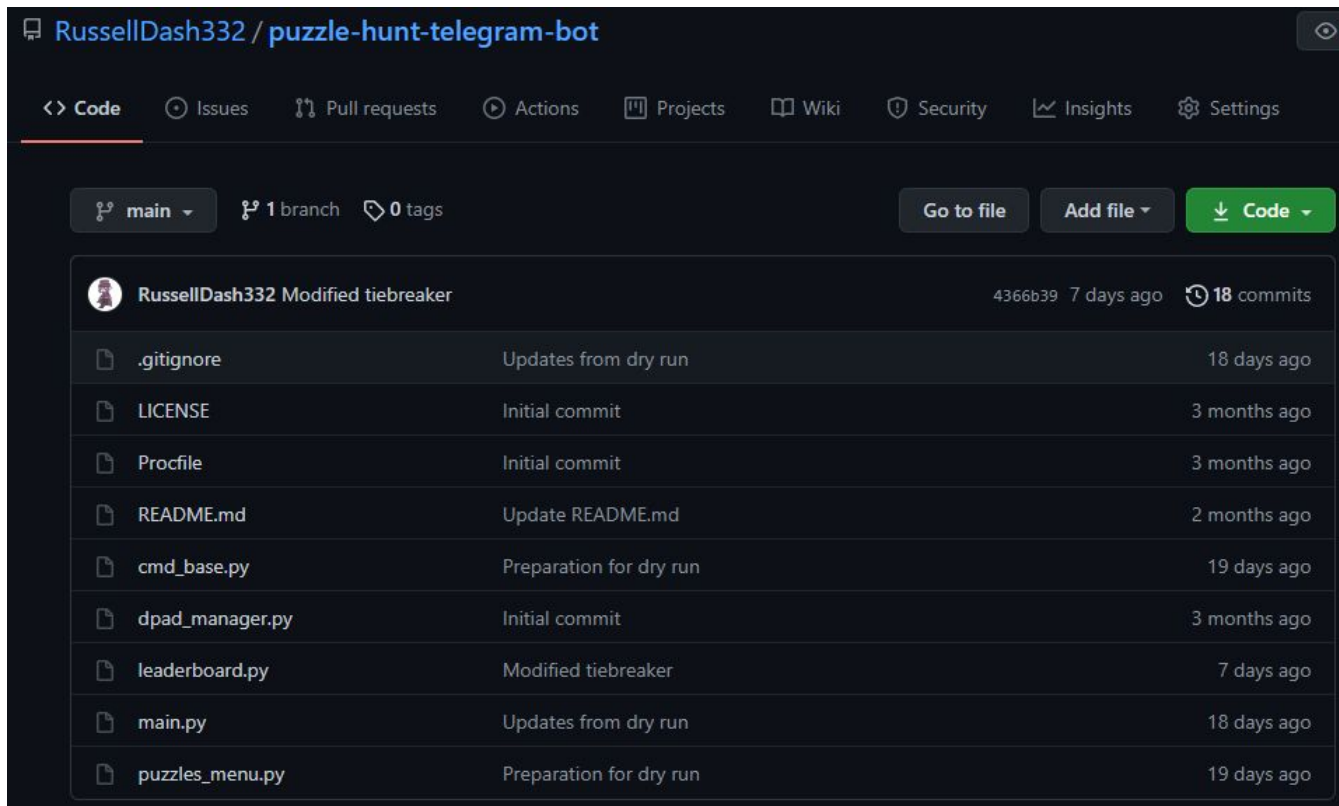
Sophomore Student at NUS

Data Science and Analytics

<https://github.com/RussellDash332>



Puzzle Hunt Telegram Bot



The screenshot shows the GitHub interface for the repository 'RussellDash332 / puzzle-hunt-telegram-bot'. The repository is on the 'main' branch, has 1 branch, and 0 tags. The file list includes:

File	Description	Commit Date
.gitignore	Updates from dry run	18 days ago
LICENSE	Initial commit	3 months ago
Profile	Initial commit	3 months ago
README.md	Update README.md	2 months ago
cmd_base.py	Preparation for dry run	19 days ago
dpad_manager.py	Initial commit	3 months ago
leaderboard.py	Modified tiebreaker	7 days ago
main.py	Updates from dry run	18 days ago
puzzles_menu.py	Preparation for dry run	19 days ago

Kattis

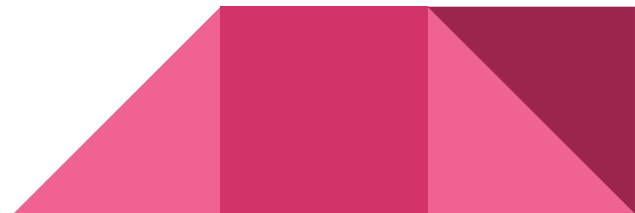
<https://open.kattis.com/>



LeetCode



HackerRank



Shopee Code League 2020



Shopee Code League 2021



Data Analytics Competition
Data Science Competition
Short Programming Contest

DSA1101: Introduction to Data Science

A must-take module as a DSA freshman in NUS!



Machine

Human

Data Management

Human Cognition

Data Mining

Perception

Machine Learning

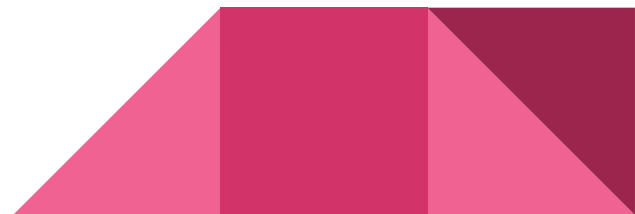
Visualization

Story Telling

Business Intelligence

Decision Making
Theory

Statistics

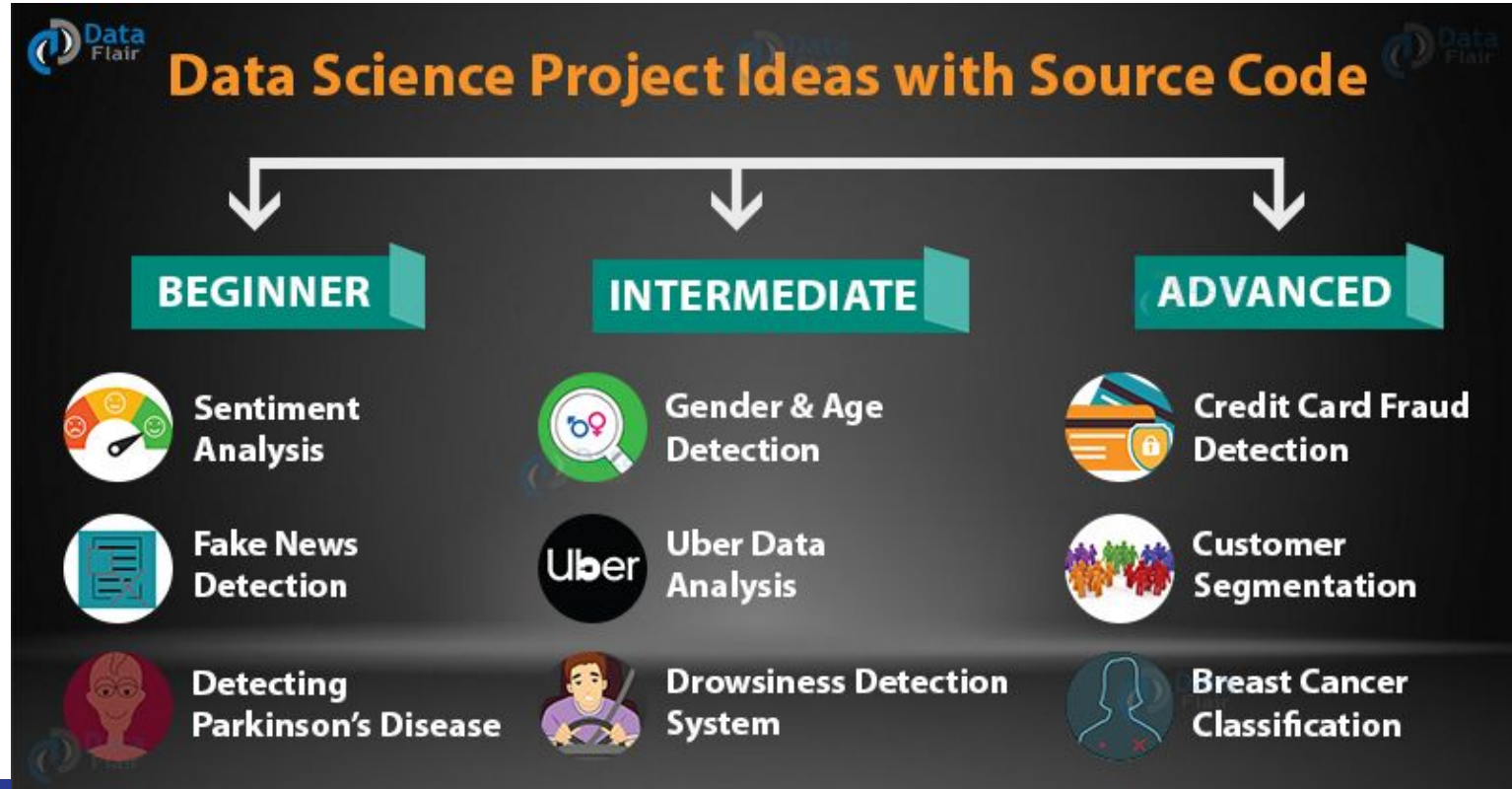


Introduction to Python with Data Science!

- Pandas -> Week 2
- Numpy -> Week 3
- Matplotlib -> Week 4
- And many more...



Finally... a data science project!



Data Analytics.. What's the difference?

In a nutshell, a more specific subset of data science.

(Shopee Code League 2020)

- The process of extracting, processing and analysing data
- Deriving insights (useful information) from data
- Using the insights from data to support decision making
- Drawing a valid conclusion



Applications

- Retail industries can collect data about customer habits using data analytics to better understand them
- Usage of AI and data analytics to improve the efficiency of clinical trials

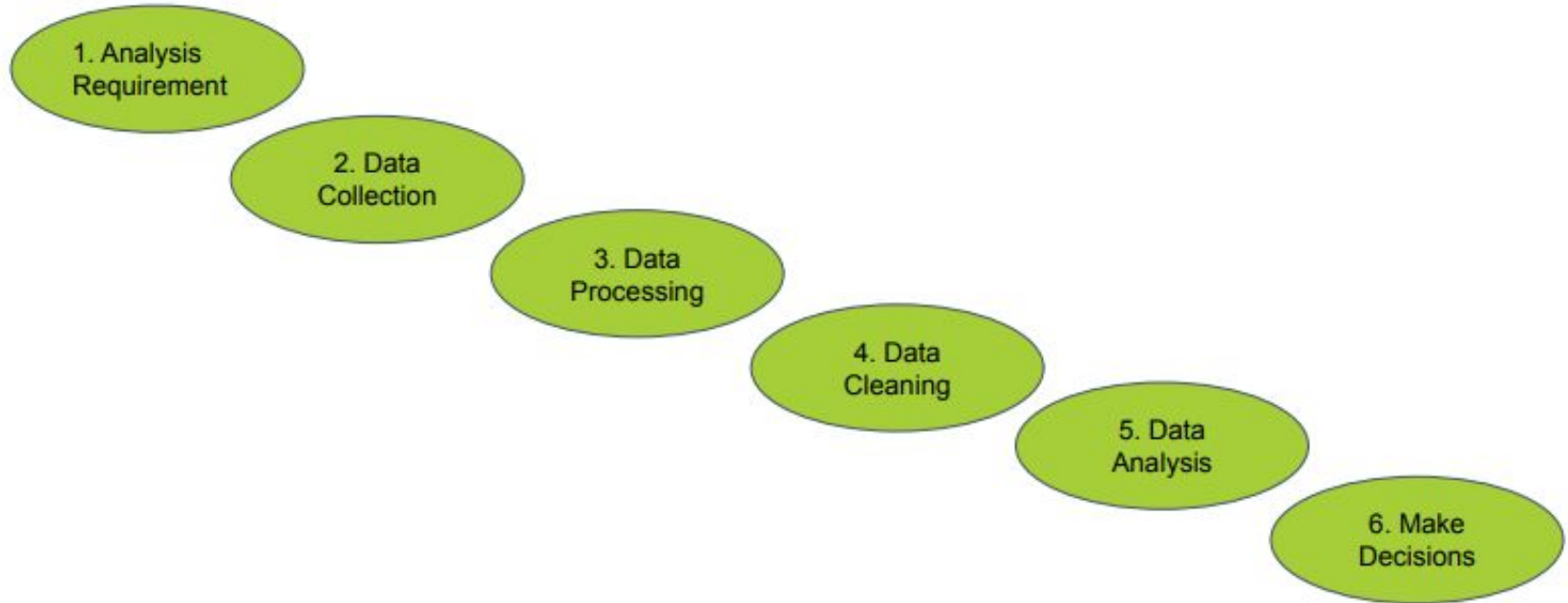
These are just two of a lot of applications!

Statistics also plays a big role, especially when conducting analysis!



Steps of Data Analytics Process

(Shopee Code League 2020)



Steps of Data Analytics Process

In this series of workshops...

- Analysis Requirement : what you need before you start conducting an analysis
- Data Collection : where is the data from? Government? Company?
- Data Processing : Pandas!
- Data Cleaning : Also Pandas!
- Data Analysis : Matplotlib!
- Make Decisions : Machine Learning



Steps of Data Analytics Process

In this series of workshops...

- Analysis Requirement : what you need before you start conducting an analysis
- Data Collection : where is the data from? Government? Company?
- Data Processing : Pandas! (+ NumPy)
- Data Cleaning : Also Pandas! (+ NumPy)
- Data Analysis : Matplotlib! (+ NumPy)
- Make Decisions : Machine Learning (+ NumPy)



Analysis Requirement

Python 3.8.6 at least

```
pip install pandas numpy matplotlib scikit-learn tensorflow
```



or...

Probably will be mostly used
during Week 4 or other workshops



Google Colab!



Google Colab

- Google Colab is a free cloud service based on Jupyter Notebooks and it supports free GPU (Graphics Processing Unit)!
- Allows you to share your code with others
 - Others can view how you wrote your code and provide feedbacks
 - Similar to Google Doc and Google Drive



Google Colab

- Often, we encounter errors on Python where it is unaware of which module you want

```
>>> import numpy as np
Traceback (most recent call last):
  File "<pyshell#0>", line 1, in <module>
    import numpy as np
ModuleNotFoundError: No module named 'numpy'
>>> |
```

Google Colab

```
Python 3.8.10 (tags/v3.8.10:3d8993a, May 3 2021, 11:48:03) [MSC v.1928 64 bit (AMD64)] on win32
```

```
Type "help", "copyright", "credits" or "license()" for more information.
```

```
>>> import tensorflow
```

```
Traceback (most recent call last):
```

```
  File "<pyshell#0>", line 1, in <module>
```

```
    import tensorflow
```

```
  File "C:\Users\LENOVO\AppData\Local\Packages\PythonSoftwareFoundation.Python.3.8_qbz5n2kfra8p0\LocalCache\local-packages\Python38\site-packages\tensorflow\__init__.py", line 41, in <module>
```

```
    from tensorflow.python.tools import module_util as _module_util
```

```
ModuleNotFoundError: No module named 'tensorflow.python'
```

Let's get started!

Sign in to your Google Account.

<https://colab.research.google.com/>



[Examples](#)[Recent](#)[Google Drive](#)[GitHub](#)[Upload](#)

Filter notebooks



Title



Overview of Colaboratory Features



Markdown Guide



Charts in Colaboratory



External data: Drive, Sheets, and Cloud Storage



Getting started with BigQuery

[New notebook](#)[Cancel](#)



Table of contents

- Getting started
- Data science
- Machine learning
- More Resources
- Machine Learning Examples
- Section

+ Code + Text Copy to Drive

Connect Editing

What is Colaboratory?

Colaboratory, or "Colab" for short, allows you to write and execute Python in your browser, with

- Zero configuration required
- Free access to GPUs
- Easy sharing

Whether you're a **student**, a **data scientist** or an **AI researcher**, Colab can make your work easier. Watch [Introduction to Colab](#) to learn more, or just get started below!

Getting started

The document you are reading is not a static web page, but an interactive environment called a **Colab notebook** that lets you write and execute code.

For example, here is a **code cell** with a short Python script that computes a value, stores it in a variable, and prints the result:

```
[ ] seconds_in_a_day = 24 * 60 * 60
seconds_in_a_day
```

[Examples](#)[Recent](#)[Google Drive](#)[GitHub](#)[Upload](#)

Filter notebooks



Title	Last opened ▼	First opened ▼	
Welcome To Colaboratory	12:27 AM	Jun 9, 2020	
Project Mushroom Edibility (Part IV).ipynb	12:18 AM	June 15	
Part2_Music_Generation_Solution.ipynb	June 18	June 18	
Part1_TensorFlow_Solution.ipynb	June 18	June 18	
Part2_Music_Generation.ipynb	June 18	June 18	

[New notebook](#)[Cancel](#)



Untitled0.ipynb ☆

File Edit View Insert Runtime Tools Help

Comment

Share



+ Code + Text

Connect ▾

Editing





+ Code + Text

✓ RAM
Disk

Editing



```
[1] a = 1 + 2  
    b = "Hello world!"
```

Ini adalah sebuah teks

```
[2] print(a)  
    print(b)
```

```
3  
Hello world!
```

```
[4] a  
    b  
  
'Hello world!'
```

```
▶ import pandas as pd  
import numpy as np  
import matplotlib.pyplot as plt  
  
import tensorflow as tf
```

More to learn and practice

kaggle

towards
data science

