

# Data Science Knowledge Repository

## Contents

- Data analysis process
  - CRISP DM
- Defining or understanding a problem
  - Asking Specific, Measurable, Actionable, Relevant, Timely (SMART) questions
  - Communicating with stakeholders
    - Who is the audience
    - What do they already know
    - What do they need to know
    - How to best communicate what they need to know
- Data collection
  - Web scraping with Python (BeautifulSoup, Selenium)
- Git / Github
  - Working with Git
  - Working with Github
- Working with Excel
  - Formulas
  - Functions
  - Pivot tables
  - Data validation
  - Data visualization with Excel
  - Misc. - Google Sheets shortcuts
- Working with SQL
  - SQL Basics
  - Data Definition Language (DDL)
    - Used for defining objects (tables) - define, change, or drop data
  - Data Manipulation Language (DML)
    - Used for manipulating data in tables - read and modify data
    - Create, Read, Update, Delete rows (CRUD) operations
  - Window Functions
  - Accessing a database with Python
    - Handling sensitive info with environment variables
    - Showing query results in Pandas format
- Data Visualization
  - Tableau
  - PowerBI
- Working with Jupyter notebook/lab
  - Creating a desktop app and configurations
  - Writing markdown files
  - Markdown shortcuts
- Working with Python
  - Data wrangling
  - Exploratory data analysis (EDA)
- Practical statistics for data science
- Machine learning
- Presentations
  - Powerpoint

# Data Analysis Process

Pass

## Defining or Understanding a Problem

Pass

## Data Collection

Pass

## Git / Github

Contents

- Git commands
- Github
  - Setting up github profile readme
  - Creating a project repository
  - Version Control
  - Commits
  - Push
  - Pulls
  - Merges

## Git Commands

References:

- [Luke Barousse](#)
- [Ken Jee](#)

## GitHub

Github profile readme.md configurations

References:

- [Yu Shi](#)
- [Emoji cheatsheet](#)

Creating a project repository

Pass

Version Control

Pass

Commits

pass

Push

pass

Pulls

pass

Merges

pass

## Working with Excel

### Contents

- Formulas
- Functions
- Pivot tables
- Data validation
- Data visualization with Excel
- Google Sheets shortcuts

### References:

Luke Barousse

[Excel Skills for Business](#)

[Edureka](#)

## Formulas

## Functions

## Pivot Tables

## Data Validation

## Data Visualization with Excel

## Google Sheet shortcuts (Windows)

Bullets	CTRL + SHIFT + 8
Headings (1-6)	CTRL + ALT + [number (1-6)]
Hyperlinks	CTRL + K
Indent Left	CTRL + [ (left bracket)
Indent Right	CTRL + ] (right bracket)
Normal Text	CTRL + ALT + 0
Number List	CTRL + SHIFT + 7
Strikethrough	ALT + SHIFT + 5
Superscript	CTRL + . (period)
Subscript	CTRL + , (comma)

# Working with SQL

## Contents

- SQL Basics
- Data Definition Language (DDL)
  - Used for defining objects (tables) - define, change, or drop data
  - CREATE, ALTER, TRUNCATE, DROP
- Data Manipulation Language (DML)
  - Used for manipulating data in tables - read and modify data
  - Create, Read, Update, Delete rows (CRUD) operations
  - INSERT, SELECT, UPDATE, DELETE
- Window Functions
- Accessing a database with Python
  - Handling sensitive info with environment variables
  - Showing query results in Pandas format

## References:

- [Luke Barousse](#)
  - [SQL for Data Science](#)
- [SQLite Window Functions](#)

# Data Visualization

Pass

# Working with Jupyter Notebook / Lab

Pass

# Working with Python

Pass

# Practical Statistics for Data Science

## Reference:

- [Shashank Kalanithi](#)

# Machine Learning

## Reference:

- [Shashank Kalanithi](#)

# Presentations

pass