

Contents

1	README	1
2	1 Monday, 21-Aug	1
3	2 Wednesday, 23-Aug	2
4	3 Friday, 25-Aug	2
	4.1 Week review	3
	4.2 Week preview	3
5	4 Monday, 28-Aug	3
6	Wednesday, 30-Aug	4
7	Friday, 1-Sep	4

1 README

- Notes usually based on whiteboard screenshots (GDrive).
- Includes further reading, watching and comments

2 1 Monday, 21-Aug

- R vs. Python: two dominant data science languages (more on the DataCamp blog)
- OOP = Object-Oriented Programming. See DataCamp course chapter.
- TIOBE: software quality control, also ranking site based on searches (tiobe.com/index). Little change among the top 10 languages. "TIOBE" stands for "The Importance Of Being Earnest" (play by Oscar Wilde).
- HTML = HyperText Markup Language. Not a programming language at all, but a language of meta data for layout. (learn it @freeCodeCamp).
- C++ is a superset of C, i.e. C is fully contained within C++ (more on C vs. C++ @freeCodeCamp).

- Scrum: agile project management method originating from software engineering and now used much elsewhere (Scrum in 20 min - video).

3 2 Wednesday, 23-Aug

- Opened DataCamp workspace to a sample notebook at tinyurl.com/WorkspaceDemoPython
- Workspace has sample data, code along webinars. It runs on a virtual Linux server (16GB RAM and 8 vCPUs).
- Interactive notebooks are much easier to use than files because documentation, code and output are all integrated already.
- These are the steps that we save when using workspace: 1) Opening Jupyter Lab launcher; 2) Opening a file editor for our first Python file ("hello world"); 3) saving the file with the ending .py; 4) opening a terminal and finding the file; 5) executing the file with the `python3` command.
- CSV files (.csv) are comma-separated-value files i.e. individual values for the cells of a table are separated by a comma.

4 3 Friday, 25-Aug

- "Dynamically typed" languages take care of memory management for data types (e.g. integer, floating point, character).
- "Cheat sheets" are collections of useful commands.
- Python libraries (like `pandas`) contain useful functions that you can use in your code if you link (aka `import`) the library first.
- `pandas` is a library for the manipulation of tabular data (called data frames). It is usually imported as the alias `pd`.
- The interactive notebook (aka workspace) contains text cells and code cells. Text cells can be edited, code cells can be run. The output appears below the code cell.

4.1 Week review

1. What are the graded course activities?
2. What is a sprint review?
3. Which tools are we using in this course?
4. Is "Python code is indented" a programming paradigm?
5. How many paradigms does Python support?
6. Can you give an example of "modular code"?
7. Which tools can you find in DataCamp workspace?
8. Terminal / shell / Command line interface
9. What are unicorn companies?
10. `pandas` and `pd.read_csv`

4.2 Week preview

- ☐ Complete DataCamp workspace demo online
- ☐ Friday: talk about the project/first sprint review
- ☐ Tuesday: complete the first test (max 30 min)
- ☐ Tuesday/Thursday: complete the next DataCamp lesson (max 30 min)

5 4 Monday, 28-Aug

- Testing the AI coding assistant.
- Reading error messages (beginner: jump to the end).
- Positional function arguments (function decides based on position what it is).
- To get help for the `unique` function in `pandas`, use `help(unique)`.
- Syntax highlighting will help you distinguish (reserved) keywords.
- When looking for help in the documentation, jump to the end to see the examples.

- If variables or functions aren't defined, you may have to run previous cells again (see **Run** tab in workspace).
- Extracting a data frame column with the index operator `[]`, e.g. `df['Category']` for the column named `'Category'` in `df`.
- To print items one per line (with a dummy variable), loop over the values and **print** them one by one, or use a *list comprehension*.
- Python can be written in a very condensed way ("one-liners").

6 Wednesday, 30-Aug

- The dot operator allows you to access functions or attributes in so-called classes (categories), e.g. `pd.unique` or `str.lower().title()`.
- Python errors are named.
- The `[]` operator is an index operator when it follows a structure like a data frame, or the delimiter for the `list` data structure.
- Grouping records by a values of a column using `group_by`.
- The 'not defined' error message may mean that some cells (with definitions, e.g. to create a data frame) were not run.
- Looking information up online.
- Shared editing allows you to write code or text in other people's notebooks.
- Computing the `size` of a subset and sorting it.
- Quick barplot plotting with `plotly.express.bar`.
- Checking a data type with `type`.

7 Friday, 1-Sep