

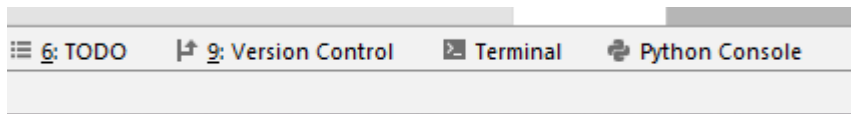
# Recommender Systems

Prof. Dietmar Jannach  
Alpen-Adria-Universität Klagenfurt, Summer Term 2021

## Assignment 1

### Task 1) Installing and running Python

- Download and install the latest version of the community edition of the *PyCharm IDE* with *Python version 3.8* (<https://www.jetbrains.com/pycharm/>).<sup>1</sup>
- Open a terminal window in the IDE, use the button on the lower end of the screen.



Type “python” to start the interactive interpreter. Type “print(‘Hello World’)” in the terminal.

```
Python 3.6.3 [Anaconda custom (64-bit)] (default, Oct 15 2017, 03:27:45) [MSC v.1900 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license" for more information.
>>> print('hello world')
hello world
>>> █
```

- Create a file called “hello.py” and type “print(‘Hello World’)”. Run the file.

### Task 2) Basic language concepts<sup>2</sup>

#### Task 2.1) Opening files<sup>3</sup>, list data structure, loops

Download the “MovieLens - recommended for education and development - Small<sup>4</sup>” dataset from <https://grouplens.org/datasets/movielens/>.

Write a program that determines the mean rating in the dataset in the following way.

- Create a list of data type “float” to store all ratings in memory.
- Open the file “ratings.csv” and read the contents line by line.
- Store each rating in the list.
- Close the file.
- Iterate through the resulting list, sum up the values and calculate the average at the end.
- Print the result.

#### Task 2.2) Functions and error handling

---

<sup>1</sup> You can use <https://www.instructables.com/id/Python-Hello-World/> for Python Hello World in PyCharm.

<sup>2</sup> You can use <https://www.w3schools.com/python/> as a starting point.

<sup>3</sup> You can use [https://python4mpia.github.io/pure\\_python/files.html](https://python4mpia.github.io/pure_python/files.html) for file handling in plain Python.

<sup>4</sup> <http://files.grouplens.org/datasets/movielens/ml-latest-small.zip>

All calculations from Task 2.1 should now be done within a function called “computeMeanRating”, which takes a file name as an input and returns a float as a result.

Define this function, implement appropriate error handling procedures (including exception handling in case the file cannot be read or found), and write a main function that invokes the method.

### **Task 2.3) Functions and return values**

Extend the function from Task 2.2 so that it returns not only the mean value, but also the mode and the median<sup>5</sup>. Write a corresponding test method.

### **Task 2.4) More data structures and file handling**

Our next goal is to analyze the genres that are appearing in the file “movies.csv”.

Write a procedure that takes the file name as a parameter and prints the following on the screen:

- All distinct genre names that appear in the file. You can use the Python `csv` module.
- For each genre, determine to how many movies it was assigned. Use a dictionary (genre -> counter) to save the number of genre assignments.
  - Print the number of movies per genre
  - Determine and print out the most popular genre.
- Optional: Sort the genres by the number of movies they are assigned to in descending order. Use a suitable library function.

### **Task 2.5) Modules and classes**

Define a Python module “utilityModule” including a class “Statistics” and add the function defined in Task 2.2 as a method to this class.

Write a test program that invokes the method (and thus prints the mean rating in the dataset).

---

<sup>5</sup> Must be done in plain Python, without using any libraries.