# Project

# Important dates

- Monday 30 September: Present Phase 1 & 2
- Monday 7 Oct: Final presentation & documentation
- Thursday 10 October: Final presentation & documentation
- Friday 11 October: Optional test: R / Python (whichever one you need to write to improve your mark)

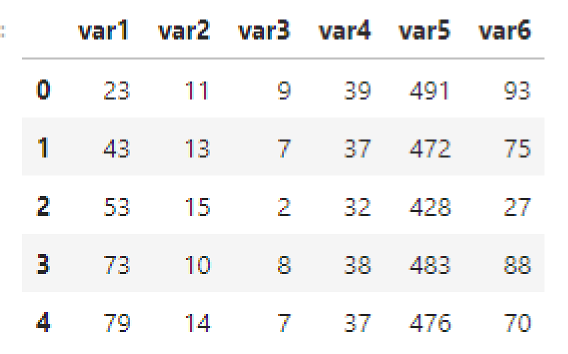
# Prepare for phase 1

- Read the data
- Create a new data frame containing only the rows about your group’s sport
- From this data frame, create 2 data frames:
  - Set 1: One containing only the columns “perform” and all the continuous x-variables
  - Set 2: Same as above, but also include “gender”
- Use Set 1 to obtain correlations
- Identify the variables that show quite a significant correlation with “perform”
- Use “perform”, “gender” and these identified variables to create summary statistics and graphs.

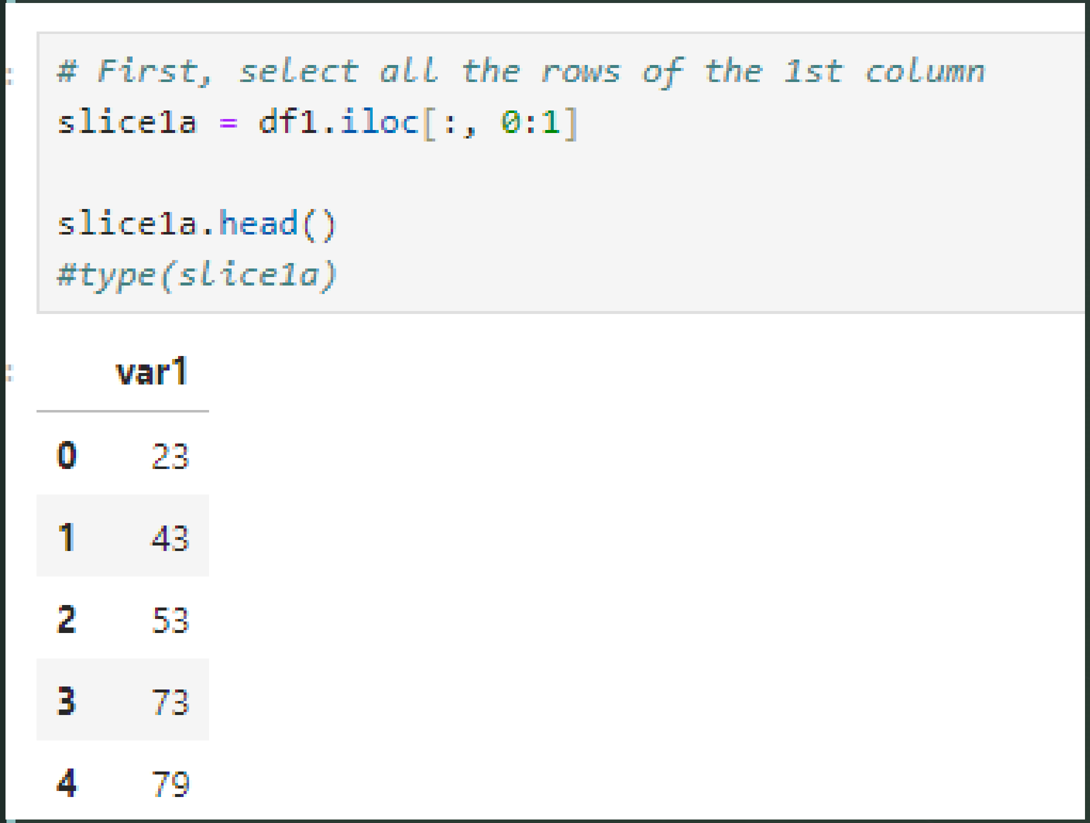
# Selecting multiple columns from a large data frame Example

- Create a “slice” of the data frame (subset) that contains all the rows of the 1st and 3rd to 6th variables.

- Data provided:



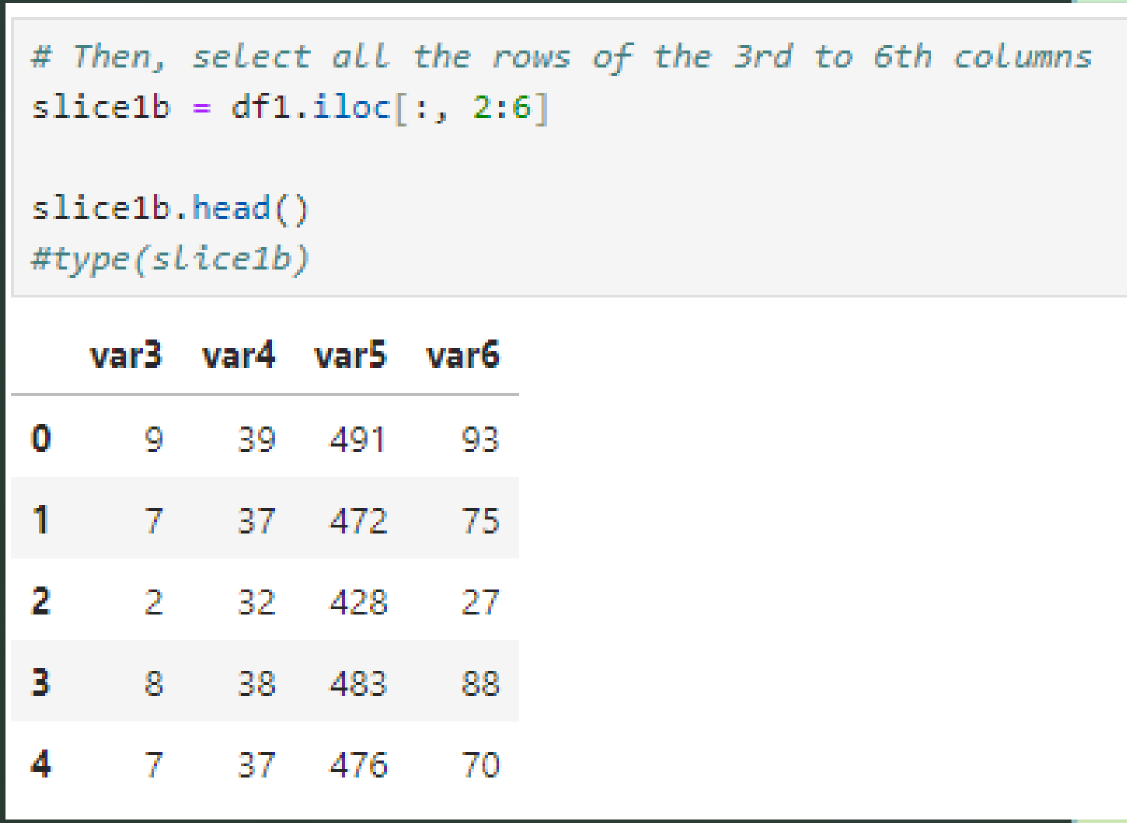
|   | var1 | var2 | var3 | var4 | var5 | var6 |
|---|---|---|---|---|---|---|
| **0** | 23 | 11 | 9 | 39 | 491 | 93 |
| **1** | 43 | 13 | 7 | 37 | 472 | 75 |
| **2** | 53 | 15 | 2 | 32 | 428 | 27 |
| **3** | 73 | 10 | 8 | 38 | 483 | 88 |
| **4** | 79 | 14 | 7 | 37 | 476 | 70 |



```python
# First, select all the rows of the 1st column
slice1a = df1.iloc[:, 0:1]

slice1a.head()
#type(slice1a)
```

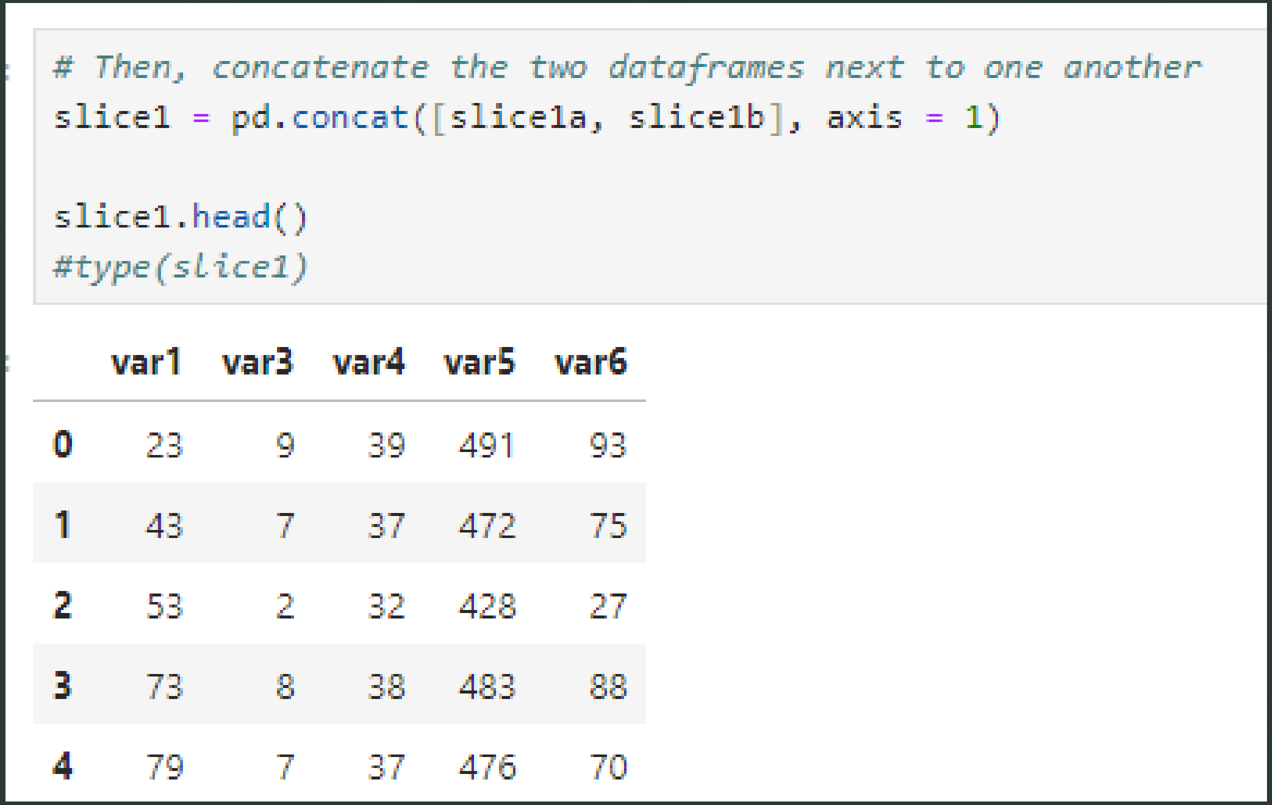
|   | var1 |
|---|---|
| **0** | 23 |
| **1** | 43 |
| **2** | 53 |
| **3** | 73 |
| **4** | 79 |



```python
# Then, select all the rows of the 3rd to 6th columns
slice1b = df1.iloc[:, 2:6]

slice1b.head()
#type(slice1b)
```

|   | var3 | var4 | var5 | var6 |
|---|---|---|---|---|
| **0** | 9 | 39 | 491 | 93 |
| **1** | 7 | 37 | 472 | 75 |
| **2** | 2 | 32 | 428 | 27 |
| **3** | 8 | 38 | 483 | 88 |
| **4** | 7 | 37 | 476 | 70 |

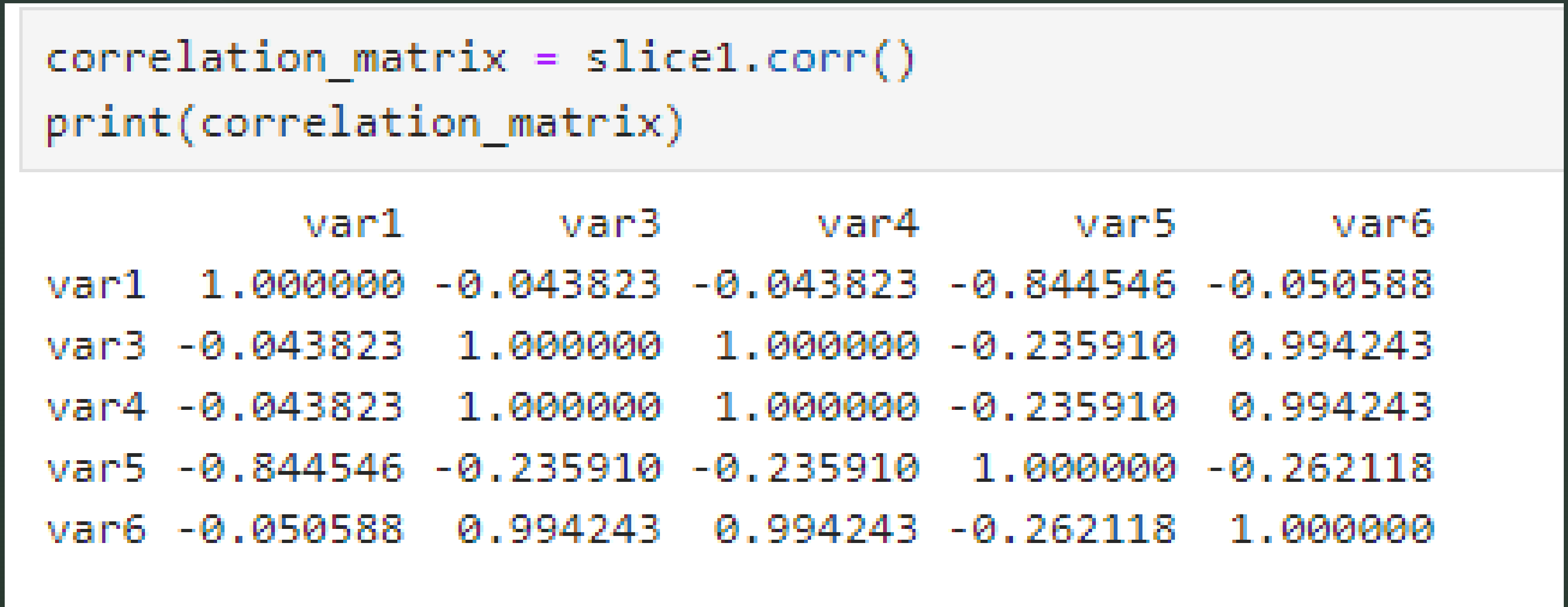


```python
# Then, concatenate the two dataframes next to one another
slice1 = pd.concat([slice1a, slice1b], axis = 1)

slice1.head()
#type(slice1)
```

|   | var1 | var3 | var4 | var5 | var6 |
|---|---|---|---|---|---|
| **0** | 23 | 9 | 39 | 491 | 93 |
| **1** | 43 | 7 | 37 | 472 | 75 |
| **2** | 53 | 2 | 32 | 428 | 27 |
| **3** | 73 | 8 | 38 | 483 | 88 |
| **4** | 79 | 7 | 37 | 476 | 70 |

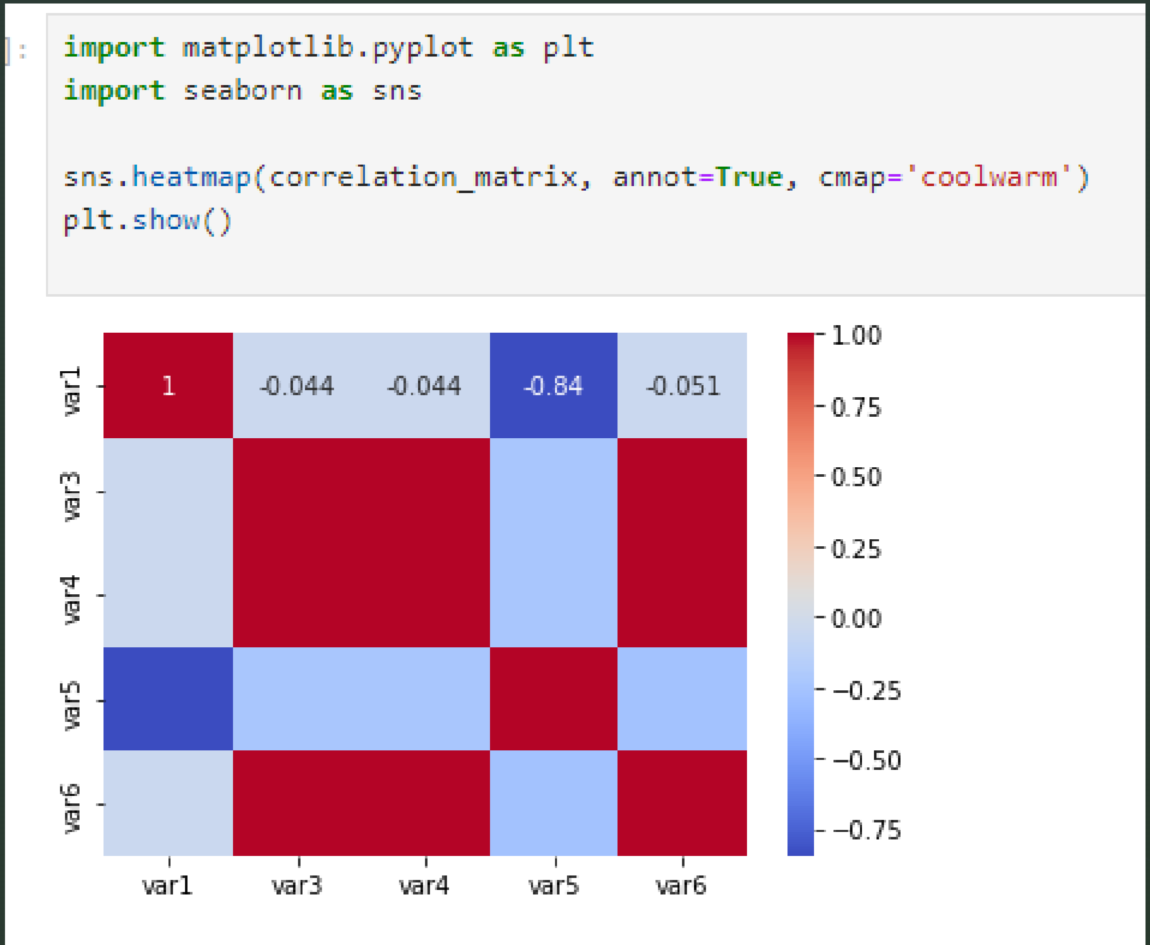
# Calculate correlation matrix



```
correlation_matrix = slice1.corr()
print(correlation_matrix)
```

```
          var1      var3      var4      var5      var6
var1  1.000000 -0.043823 -0.043823 -0.844546 -0.050588
var3 -0.043823  1.000000  1.000000 -0.235910  0.994243
var4 -0.043823  1.000000  1.000000 -0.235910  0.994243
var5 -0.844546 -0.235910 -0.235910  1.000000 -0.262118
var6 -0.050588  0.994243  0.994243 -0.262118  1.000000
```

# Create a heatmap from the correlations



```
import matplotlib.pyplot as plt
import seaborn as sns

sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
plt.show()
```

# What is required from you

- The team must decide who will do what.
- Every person needs to create 3 different graphs, using the “Perform”, “Gender” and fairly significant variables found from the correlation analysis.
- Think which hypotheses you would like to explore using graphs, e.g.:
  - What is the effect of gender on Var X?
  - Is Var X normally distributed?
  - Is there a relationship between Var X and Var Y ?
- Create nice graphs, adding titles, labels, legends, colours, lines, etc.

# How will it work?

- **Each team member** is responsible for creating the following outputs for this project:
  - Phase 1: Descriptive stats & graphs
  - Phase 2: Hypothesis tests
  - Phase 3: Model(s)
  - Phase 4: Literature study
- What you create, you need to put in a Powerpoint and you need to present YOUR section.
- You will receive an INDIVIDUAL mark.
- **As a team**, you will coordinate your efforts so that everyone does something unique for the understanding of the sport you are responsible for.
- E.g. Person 1 does histogram to test normality of X1, person 2 does a scatterplot between X1 and X2, person 3 does a boxplot of X3 versus gender. They all work on the ice skating dataset, but each contributes a different part for each phase.

# Presentation 1

- Time limit: 4 minutes per student
- Each student will present his/her work on phases 1 & 2 using professional-looking Powerpoint slides. You can be informally dressed (as you usually come to class).
- If you cannot attend class on the day, you have to arrange with me BEFORE the time. Then you need to create a video with a voice-over and submit that.
- Ensure that your code is neatly structured and have suitable comments.
- Submit your Powerpoint slides.
- Submit your code in Word format as well as Python format.

# Presentation 2

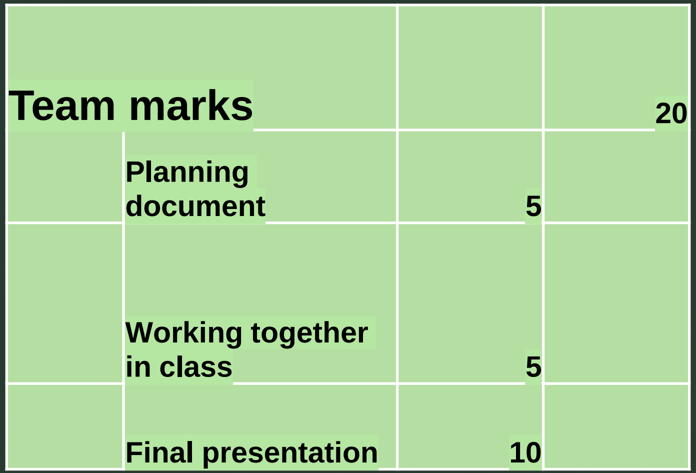
- Time limit per student: 5 minutes
- Each student will present his/her work on phases 3 & 4 using professional-looking Powerpoint slides. You may add a summary of the work done in phase 1 & 2 to support the work of phase 3&4, but don’t go over the time!! You can dress slightly more formally (but please, don’t spend money!).
- This counts as a Python test, therefore, you must present in person.
- Ensure that your code is neatly structured and have suitable comments.
- Submit your Powerpoint slides.
- Submit your code in Word format as well as Python format.
- Complete a questionnaire on your experiences with this project.

# Rubric
# Individual marks

|  |  |  |  |
|---|---|---|---|
| Slides: |  |  | 12 |
|  | Grammar and spelling | 3 |  |
|  | Visually pleasing | 6 |  |
|  | Font size colour readible | 3 |  |
|  |  |  |  |
| Presentation: |  |  | 13 |
|  | Audible | 3 |  |
|  | Makes eye contact | 3 |  |
|  | Confidence | 5 |  |
|  | Sticks to time | 2 |  |
|  |  |  |  |
| Graphs: |  |  | 15 |
|  | Effort | 8 |  |
|  | Interpretations | 7 |  |

|  |  |  |  |
|---|---|---|---|
| Hypotheses: |  |  | 15 |
|  | Effort | 8 |  |
|  | Interpretations | 7 |  |
|  |  |  |  |
| Model(s): |  |  | 15 |
|  | Effort | 8 |  |
|  | Interpretations | 7 |  |
|  |  |  |  |
| Coding: |  |  | 4 |
|  | Structure | 2 |  |
|  | Comments | 2 |  |
|  |  |  |  |
| Literature: |  |  | 6 |
|  | Effort | 2 |  |
|  | Referencing | 2 |  |
|  | Writing style | 2 |  |

# Rubric
# Team marks



| | | | |
|---|---|---|---|
| **Team marks** | | | 20 |
| | Planning document | 5 | |
| | Working together in class | 5 | |
| | Final presentation | 10 | |