## First data analytics project

Insert the expected completion date below:

|  |
| --- |
| 21.05.2022 |

* The goal is to go through this list at least 3 times, each time increasing the number of questions to answer by 3. 😊
* Feel free to google, watch some tutorials or read some blogs if you encounter any uncertainty. But please keep in mind the end date, like you would in a professional project ^^. Then apply it to your solution.

|  |  |
| --- | --- |
| Steps – Time schedule | End Date |
| * Write down 5 measurable things, that are interesting for you. | **18.05.2022** |
| * Research and find data related to the chosen topics.   Try to google it including “dataset” query, use known sites with data or use your own scrapers to collect it (advanced).   * Choose the topic where a lot of data is available. | **18.05.2022** |
| * Put yourself in the role of manager / owner / politician / coach and come up with **3 different questions** that you will be able to answer after examining the collected data. Those questions should use your business intelligence in your chosen topic 😊 | **18.05.2022** |
| * Install database of your choice and add your data to it. | **18.05.2022** |
| * Answered question 1: Most points? | **19.05.2022** |
| * Answered question 2: Least injuries? | **19.05.2022** |
| * Answered question 3: Player costs? | **19.05.2022** |
| * Using SQL or your chosen technology create a view with answers. | **19.05.2022** |
| * Install visualisation tool to illustrate your answers. | **20.05.2022** |
| * Connect your database with your visualisation tool. | **20.05.2022** |
| * Download a few dashboard examples / templates for your visualisation tool. (to make it easier try to use ready solutions) | **20.05.2022** |
| * Prepare a report including your answers and visualised data supporting it. | **21.05.2022** |
| * Export your report as PDF and send it to me 😉 | **21.05.2022** |
| * Add your SQL code to GitHub repository. Be proud of your work! | **21.05.2022** |

1. **Write down 5 measurable things, that are interesting for you:**

* NBA playoffs / teams / players

https://www.kaggle.com/datasets/martinellis/nhl-game-data

* ~~Retail / Amazon~~

~~https://jmcauley.ucsd.edu/data/amazon/~~

* ~~Small electronics~~

~~https://www.kaggle.com/datasets/mkechinov/ecommerce-purchase-history-from-electronics-store~~

* ~~Programming~~
* ~~Stock market~~

1. **I’ve found many different datasets and chosen: …**

|  |
| --- |
| NHL League – It was the easiest and largest dataset to find! |

1. **Let’s assume that I’m a: …**

* COACH

**, and I want answers to the following questions:**

**(First iteration)**

* Which player scored the most points in last season? How many points each of them scored in the last season?
* Which player had the least injuries in last season?
* How much do those players cost? What is they price?

**(Second iteration)**

* …
* …
* …

**(Third iteration)**

* …
* …
* …

1. **I’ve created a view and I've chosen visualisation framework: …**

* Metabase.

After trying with Grafana (which is timeline oriented) I switched to opensource solution that allows more than just showing the change of parameters in time.

1. **I’ve chosen the following report template:**

* … I needed to create my own cause Metabase doesn’t provide full templates. The effects are presented in pdf report file 😉

1. **I've created a report and send it over to @rojberr 😉**

****