The DreamTeam dataset

- **Platform for e-sports** (only CS:GO at the moment)

- **USERS**
  - Profile information (id, gender, nationality, language, etc)
  - “Follow” relationships
  - Reviews
  - Performance (rounds played, wins, “kills”, etc)
  - Ratings / Self-rates
  - Up-votes (liking a review)
  - Roles

- **TEAMS**
  - General information (id, country, language, social network links, etc)
  - Invites / requests
  - Vacancies

- **Presentation + Q&A with DreamTeam on January 25**

- **Contact: g.livan@ucl.ac.uk (Giacomo Livan)**

1st assignment - due February 16, 2018

- **Download** the DreamTeam dataset: https://www.dropbox.com/s/m2nmv0owjjbeif2/prod_dump_Jan9.tar.lzma?dl=0

- **Import** the dataset with SQL software

- Provide an **overall description** of the dataset. Explain its structure, provide an overview of the main variables and explain what aspects of the game / platform they capture

- Study the **distributional properties of players' activity.** Plot the distribution of at least three variables related to activity (e.g., total shots hit, total kills, time spent playing, etc) and study their distributional properties: measure the first four moments, discuss which class of distribution might be best suited to describe the data and **why,** and **fit it to the data.**

- Describe the methodologies employed to obtain the results of your project. **If you use off-the-shelf software, you are expected to outline how they work and which methods they use.**

- Summarize your findings in an **individual report** of less than 2000 words

- **Marking criteria**
    - A **clear structure** with introduction, main body / analysis, and conclusions
    - **Readable plots**, with captions, axis labels, etc
    - **Clarity of language**, i.e. explain things for a reader unfamiliar with the game and the platform
    - Clear **explanation of the methodologies** used

- This will account for 15% of your final mark

2nd assignment - due March 22, 2018

- Choose a **statistical problem** / research question for your final report and tackle it
- This can be a research question of your choice, or one of the challenges suggested by DreamTeam (see list of “Data Science tasks” by DreamTeam)
- Present the main results in a **1-page individual executive summary** containing one figure or one table. Avoid excessive technicalities, which can be explained extensively in the final report.
- Marking criteria:
  - **Clarity** of language and exposition
  - **Self-consistency** of the summary
  - **Readability** of the figure or table that you include (it is mandatory to include one of the two)
  - Ability to identify and summarize your key findings findings and make them easy to grasp
- This will count as 15% of your final mark

Final report - due March 29, 2018

- Combine a presentation of the DreamTeam dataset, the results of your 1st assignment and a detailed description of your project in an **individual final report of no more than 5000 words.**

- Feel free to re-use the material from your 1st assignment report or to change it / improve it as you like based on the feedback you received

- Describe the methodologies employed to obtain the results of your project. **If you use off-the-shelf software or packages, you are expected to outline how they work and which methods they use.**

- Marking criteria:
  - Is the **introduction** clearly reporting the essential points?
  - Are the **methodologies** explained well?
  - Are results and finding clear and correct? (Keep in mind: **negative results are OK!**)
  - Are figures, tables, and captions **comprehensible**?
  - Are **conclusions** reporting the **main findings** and highlighting the **limitations** and possible **extensions** of your work?

- The final report will count as 40% of your final mark

Data science tasks

Challenge	Goal, Details	Examples, use cases
General, Classification		
Data Analysis, any hidden data correlations and dependencies	Discover interesting facts, hidden correlations and dependencies for both internal purposes and to present catching infographics for platform users as a result motivate them to use the product and make it more attractive and valuable	• Average rating per country • Correlation of any metrics, e.g. in Brazil snipers are much stronger • How quick players per different skill group find their teams
Players and teams clustering, where a player-team among other players-teams	To present to the audience the structured and definitive look at the platform audience and to show where a player / team is on their way to excellence compared to others	For example, one set of possible clusters, let's say per region: • Newbie • Semi-professional • Professional Possible use case could be to show that you are a Newbie, but high chances to move to Semi-professionals in your region in next month
Tips how to progress (boost skills) for player / team	Particular tips what to do to boost their skills	For example, based on statistics from Steam and model studied on most successful players and teams progression, suggest tips on what to do to speed up the growth. For instance, player should spend 10% more time in game weekly; Team should try planting the bomb more often on Dust II in the second third of the match.
Best players-teams	Smart algorithm of defining best players and teams, as a result increase user engagement and motivation to growth	You are in 5% best players on the platform Your team is in 10% best teams on the platform Leaderboards, ranking by gun, map etc. Again could be per region
Player-team with the best potential	Identify talents on early stages, as a result teams will work actively to hire such people, develop them	Suggest best talents for teams
Most progressing players-teams	Again highlight most attractive players for teams	Suggest best talents for teams
Recommended auto rating for players, auto rating for teams	Smart automation	
Auto definition of player roles based on their skills	Smart automation	
Recommendations	Enrich platform functionality, simplify users experience, speed up search of needed players/teams	

Player for team (for invites flow)	see above	When Team Owner searches for free players system returns most relevant results, recommended for particular team either on the top of the search or on a separate section. Another case when system analyses player roles and team needs. As a result system recommends players with specific roles and skills to particular teams in order to boost teams by covering some specific areas, for example, empower the team by good sniper
Player for team vacancy	see above	When Team Owner creates vacancy system suggests most efficient players and Team Owner might sent an invite.
Team vacancy for player	see above	System sends notification to free payers with summary of best suitable vacancies.
Team for team for practice games	see above	Team is being notified with contacts of relevant peers to have practice games for training purposes
Social graph	see above	Sections like on social networks, recommending teams/players of mutual skills, game tactics, interests etc.
Predictions	Increase engagement of users to the platform, increase value	
Chances to win: team vs another team	Looks as a really challenging task especially taking into account actual information on the platform. An idea is to search for additional info on other platforms and Steam API to be able to teach the model based on game statistics and then apply the same model for the platform	For practice games matchmaking, in future for tournaments
Prediction when player-team achieves next skill level if proceeds with the same constant performance	motivate player	
Probability that Player will be accepted to join the particular vacancy in a team	Shorten time of filling in the position, more relevant candidates, higher satisfaction	