

Certification Project for the Data Science Training Program

Description: Solve a data problem of your own choice. There are no limitations for the final projects and datasets, you are free to choose. Maybe avoid using the datasets that were presented during the program :)

It is beneficial to start your own project as soon as possible; that way, you can play with data, apply the learnt concepts before you forget them, and ask questions from the instructors during the labs and on the community platform.

You are free to work in a group of 2-3 people or individually.

Apply all concepts learnt in the course

- Consult the first session, there is everything mentioned briefly
- Include (try to include) all concepts learnt and relevant for your problem, and justify why some of the methods were not applied

List of the concepts we will go through:

- Descriptive Analysis
- Visualization
- Data Manipulation
- Hypothesis Testing
- Data Engineering
- Modeling (Linear and logistic regression; basic feature selection; bias/variance tradeoff; regularization; diagnostics, Tree methods, Diagnostics, ensemble methods and experiment tracking, Feature Selection, Dimensionality Reduction)
- Unsupervised Modeling
- Storytelling

Presentations of the projects during the last week: **16 - 18 of December 2019**

Please ask your questions regarding the projects and the program in general at the Q&A section in the community platform and we will answer them at the lab: [link](#)

GOOD LUCK! <3