» neue fische
School and Pool for Digital Talent

Data Analytics Capstone Project

During the last 4 weeks of the Data Analytics Bootcamp, students work full time in a team on their own capstone project for putting the concepts learned in the bootcamp into practice, deepening knowledge and strengthening abilities as well as providing an example of work for their portfolio.

The capstone process is outlined here to give you a better understanding about the contents and requirements of the capstone project.

Capstone in a nutshell:

In the capstone project, students should in the end create a 10-15 minute presentation that answers a specific question about a business topic. The project involves finding the topic and iteratively developing and clarifying the question along with finding the data and performing the analytical steps required to answer the question. Students tend to work with between 2 and 4 separate data sources.

Required deliverables:

- ☐ Clearly stated business question with the background and impact explained.
- ☐ Technical data analysis completed using Python, SQL and Visualization, eg. Tableau.
- ☐ Work is stored in a github repository.
- ☐ Data is stored in a SQL database (neue fische can provide it during project time).
- ☐ At least 2 separate data sources are used.
- ☐ Final presentation of 10-15 minutes, accompanying slides are included in the github repo.

  Data reuse is granted, though your project must be genuinely yours and distinguishable from projects done before. tl:dr - no copy/paste!

Capstone project outcome effects

Group size: The capstone project is a group project with a minimum members count of 2. Groups sizes of 3 or 4 members are recommended.
Graduation Event: Small group sizes may lead to reduced visibility during the final graduation event due to more presentations need to be held during the event.
Certificates: Not fulfilling the required deliverables may lead to conclude the bootcamp with just a certificate of attendance instead of a Data Analyst certification.

Capstone and data analytics workflow steps:

Project definition: Understand stakeholder needs, potential outcomes, actions, areas of uncertainty, and success criteria plus measurement of success.
Scope out project in reverse order: Desired outcome, potential actions, decisions to be made, predictions to be used, data requirements.

Find data: Access data via provided files, databases and/or using APIs.

Clean and manipulate data: Data wrangling using taught techniques such as identifying duplicates, finding missing values, detecting outliers. Store data in SQL database.

Understand data: Exploratory data analysis to establish validity of dataset.
Finding the distribution of data, the presence of outliers and the correlation between different columns.

Generate insights: Data visualization using Python libraries and/or Tableau to highlight the distribution of data, relationships between data, and the composition and comparison of data.

Communicate insights to stakeholders: Charts and dashboard construction using Tableau or Python to highlight insights in an intuitive, appealing, and easy to understand way.

Drive decision making and action: Presentation demonstrating your ability to clarify your analysis and relay your findings in a way that is easy to understand.