Details

Use this page to upload your final project. The project consists of a written report and a slideshow. Both files must be uploaded as PDFs.

(For instructions on how to do this, please read this article on **how to save PowerPoint presentations as PDF files (https://support.office.com/en-us/article/save-powerpoint-presentations-as-pdf-files-9b5c786b-9c6e-4fe6-81f6-9372f77c47c8#OfficeVersion=Windows)** .)

As a reminder, the presentation carries a 7.5% weight towards your final grade and the report carries a 12.5% weight towards your final grade.

Your presentation must make use of up to 10 slides. You will have 8 minutes at your disposal and all team members are required to present.

The report is expected to be no less than 10 pages and no more than 20 pages, with 1.2 spacing, including figures, tables and code listings. The report must include the following sections:

Problem statement

A description of the chosen scenario with all applicable details and requirements set forth by your client.

Proposal

Details on the plan you proposed to your client and how the company will benefit from the database system and processes that you committed to implement.

Team structure and Timeline

An outline of what each member of your team was responsible for. Also, a timeline of activities highlighting the most challenging tasks.

Database Schema

A description of the designed system, complete with logic, ER diagrams, and code listings.

Analytics Applications

While you are not required to build specific applications, you must explain how you will cover the needs of your client besides designing the backend database system.

Additional Details

All projects will be made available to the class prior to the live presentations. You are expected to browse through your classmates' projects.

Also, please make sure you demo the presentation system, your audio and video prior to the live presentations.

Submission

To complete your submission,

1. Click the blue **Submit Assignment** button at the top of this page.
2. Click the **Choose File** button, and locate your submission.
3. Feel free to include a comment with your submission.
4. Finally, click the blue **Submit Assignment** button.

Detailed Requirements & Grading

Project reports and presentations are due on the day before the final class session, as such you have to upload two PDF files, one for the presentation slides and another for the project report. It is a requirement for your final Attendance and Participation (A&P) points that you review your classmates' reports and presentations prior to the class session. They will be posted on the "Team Projects" page right after the deadline.

Project Checkpoint grades do not guarantee Final Project submission grades. Final reports will be graded independently, so please make sure you include all work. Nothing from the checkpoints may be assumed to be part of the final report.

Scope

- Assume you are a team of database experts hired by a client
- Your client has a lot of archived, unstructured data that they need to have organized for easy access and analysis
- Your client also has new data coming in that need to be automatically stored and efficiently retrieved
- Your client requires two levels of access to the data:
  - Analysts: write and execute SQL code as well as access the database through Python/R
  - Managers/C-executives: high-level overview through visualizations and interactive dashboards that automatically update when new data is stored in the database

Overview

- Your team will have to propose a scenario
- There is no specific limit but small datasets or pre-defined relational datasets are not appropriate. The main requirement is that the database design should result in at least 15 tables at 3NF.
- Develop a relational schema and load data (Python/R can automate the loading process)
- Demonstrate the efficiency of the relational database by showcasing at least 10 analytical procedures that led to valuable insights for your client.
- Build interactive dashboards in Metabase

Presentation Requirements

- Eight (8) minutes of presentation (~10 slides)
- All team members must present, speak clearly and coherently
- Presentations must summarize the report:
  - State your consultant/client scenario
  - Show a sample of the original data
  - Overview and brief explanation of the database design
  - Brief explanation of the ETL process
  - Process of interacting with data (analytical procedures):
    - Why? What are the customer needs that you designed for? What are the insights that your customer is expected to benefit from with the new relational database and relevant tools that you build for them?
    - How? Did you build custom tools? (raw SQL, Python, R, Metabase, other)
  - Very brief demo of database interaction (SQL and Metabase)
- Be prepared to answer questions

Report Requirements

- Your report must be prepared in a professional style and format. Use 12pt Arial or Times New Roman font, spaced at 1.2 lines. Report should not be longer than 20 pages. With code and figures removed, there should be 5-7 pages of text. Submit as a single PDF with all relevant attachments.
- Start by describing your consultant/client scenario. Detail the reasoning behind your choice, your motivation, the research you have performed in making your decision and your initial plan of action. Explain how your work will improve decision-making for the company that hired you. What other benefits will there be?
- Include your team contract, who was responsible for each task?
- Present a sample of your data and provide links to the full dataset. Explain why you chose this dataset.
- Present your database design and describe all steps in detail. ER Diagrams must be attached as PDF, provide link to Lucidchart too. Provide SQL code for the tables.

- Present your ETL process *in detail*. It is highly advised that all code listings are properly formatted and all code is also provided as a GitHub repo or Gist (provide links). Pasting code as images is not acceptable.
- Explain **all work** for database design and ETL. Points will be taken off if you submit ERD/code without full justification, even if they are correct.
- Showcase at least 10 analytical procedures that led to valuable insights for your client. Support each one with its corresponding code listing that delivers the results. Code may be raw SQL or Python/R. Examples of such procedures are:
  - *What reseller sales have been made, and where?*
  - *How are the salespeople performing?*
  - *Which customers are likely to buy from us?*
  - *Which products do our customers buy together?*
  - *What is the sentiment of our new product?*
- This is your chance to demonstrate your proficiency in SQL coding. Queries that perform simple actions such as counts or unfiltered averages will not receive full points. Rule of thumb: a prescribed procedure is too simple if it can be performed just as easily in MS Excel on the original tabular data
- Present your plan on how your customers will interact with the database system you designed. This has to be specific. What will you implement for analysts (direct querying) and for "C" level officers (reports)? How will analysts run the code for the analytical procedures you designed? How will executives review the results? What tools/programming languages did you implement? What were the benefits of performing database actions with a programming language? Can non-technical personnel interact with the database?
- Did you plan for redundancy and performance? Should your client host on-premises or on the cloud?
- Showcase the dashboards you produced in Metabase. Paste screenshots and explain what is presented. What are the benefits of these dashboards. You can also provide links to these dashboards, if you set up Metabase in a way that it is publicly accessible. (note: you may enable public sharing in the admin settings)
- Conclusion: summarize your goals and how these were achieved. Make sure to emphasize all benefits of RDMS, ETL and analysis. What insights were made possible due to your work, how did your client benefit from your RDMS implementation?

Grading Weights

- Report/Presentation style and formatting quality, properly formatted code with links (10%)
- Consultant/client scenario justification (5%)
- Database design (30%)
- ETL (20%)
- Analytical procedures (20%)
- Analyst vs. Executive: justification of methodology and tools (5%)

- Metabase dashboards (10%)

Please note:

You should be fair with your ratings, however your answers will not be used to grade team projects, only we (your instructors) will grade your work. Answering these questions aims to keep you engaged and critical of each others work so that you may perhaps learn something new from the great work of your colleagues, so that we can all benefit from each others work.