Group 5: Healthcare

Temesgen Fekadu
Hayden Muscha
Jack Lynn
Sam Wainright

PROJECT MANAGEMENT PLAN

Explain project and define goals; the "what" and "why". What is the scope baseline?

The consensus by Group 5 is to investigate diabetes data for the project.

The goal of the project is to provide an overall holistic view of diabetes in the US. The questions posed to complete this are the following:

1. Which demographics are most likely to develop diabetes?
2. What measurable bodily attributes contribute to the development of diabetes?
3. How are blood glucose levels tracked in real time?
4. Which demographics are exhibiting higher spikes in blood glucose levels?
5. Looking at specific regions within the US, how do different lifestyles contribute to diabetes prevalence?
6. Does food scarcity contribute to diabetes incidence?

Deliverables are outlined on the Capstone rubric and will be stored in the provided GitHub repository link:

1. Exploratory questions
2. Project management plan
3. Napkin drawings and feedback for visualizations
4. Napkin drawings and feedback for dashboard
5. Repeatable ETL report
6. Dash dashboard
7. Code (if applicable)
8. Project technical report
9. Presentation slides

Data sources.

After discussing the topic, data sources were collected to reflect the goal, and questions posed. Preliminary research began on 29 Aug. All group members researched articles that gave insight into the development of diabetes. Data sources need to address prevalence in the US, parameters influencing diabetes, and a measurement of parameters in individuals diagnosed with diabetes.

Communication and check-in process.

Morning group standup at the start of the day addresses what we worked on last night and what we are trying to get finished on the current day, both as a group and individually. The last hour of the workday is used to discuss any tasks a person personally wants to complete by the start of the next day, or if they could use assistance for the next day. Communication is maintained through the designated Microsoft Teams group chat.

Schedule with tasks, subtasks, milestones.

Timeline is 3 weeks with final due date by end of day 18 August and presentation on 19 August.

Week 1, 1 Aug: tasks involve research, brainstorming, exploratory analysis, and data cleaning. By end of week 1, 5 Aug: any data being used should be cleaned and uploaded into the database.

1 Aug

- Set up of GitHub by J. Lynn. Set up of database and data lake by H. Muscha.
- Building a dashboard with Dash or Power BI with Python.

  Dashboard being built with Dash. The goal is to have a dynamic website that can be accessed for research purposes.
- Project flow diagram created by J. Lynn and formed the outline of the project plan. Formal written project plan and planner board created by S. Wainright. To be submitted in the Brightspace capstone module rather than GitHub by 5 Aug.

2 – 3 Aug

- Everything needed for Kafka.

  Cloud ETL conducted through Microsoft Azure.

  Values for extraction with provided config information are in Databricks shared folder for cohort 25.

  The transform step of producer and consumer composed by J. Lynn using “group5cgm” as topic. Producer stream is the glucometer dataset. The dataset was researched by T. Fekadu and will be used in the Kafka-to-database portion of the project. Cleaning the dataset was conducted by T. Fekadu as well as creating the data factory pipeline with trigger.
- Census and CDC datasets were researched and cleaned by H. Muscha. The glucometer meter dataset was researched and cleaned by T. Fekadu. Health parameters dataset researched and cleaned by J. Lynn.

4 – 5 Aug

- Loaded datasets into data lake and SQL database, completed 3 Aug. Jupyter notebooks used to complete ETL loaded into GitHub 4 Aug.
- Research for deep learning models by J.Lynn on 5 Aug.

  Looking into long short-term memory (LSTM) networks and their implementation as well as any other time-series methods.
- Exploratory data analysis for ML models by S. Wainright on 5 Aug.

  Get baseline values and research ML models to use.
- Set up ETL Report and Project Technical Report – S. Wainright

  ETL .docx file in deliverables folder used as a rough draft structure, to be formally written in week 3.

  Diabetes prevalence, education by state, food insecurity, income, and NHANES datasets ETL and rough ETL report – H. Muscha

Week 2, 8 Aug: tasks are focused on planning and building. At the end of week 2, 12 Aug: machine learning models should be built, and napkin drawings have received feedback.

8 Aug

- Building the machine learning model – S. Wainright

  Create evaluation graphics that include confusion matrix, F1 values, PR vs. ROC.

  Start grid search by EOD

- Building the base deep learning model – J. Lynn

  Determine y-value and form structure

- Napkin drawings for visualizations – H. Muscha

  Determine relevant factors to be measured.

  Receive and document feedback from 2 colleagues outside of Group 5.

- Create Python script with Dash – T. Fekadu
  Load in cgm data from database

  Receive and document feedback from 2 colleagues outside of Group 5.

9 – 10 Aug

- Finalize napkin drawings based on feedback – H. Muscha & T. Fekadu

  Document if changes were made based on feedback or explanation if there was a decision not to utilize feedback.

  Save under *Visualizations* section in the Project Technical Report rough draft.

- Create visualizations for Dashboard – T. Fekadu and H. Muscha

  Determine theme.

  Provide instruction to J. Lynn and S. Wainright for visualizations necessary for dashboard and PowerPoint.

- Asses a secondary ML model – S. Wainright

  Tune both models and test for optimization

- Train DL model – J. Lynn

  Set up parallelization in co-lab.

  Evaluate performance

11 – 12 Aug

- Create ML Visualizations – S. Wainright
- Update dashboard graphics based on group and/or cohort peer feedback – H. Muscha
- Either continuation of 9 – 10 Aug or colleague assist. Do not move on to week 3. – T. Fekadu & J.Lynn

Week 3, 15 Aug tasks focus on finalizing and refining models, composition of reports, and presentation.

15 – 16 Aug

- Refine DL model – J. Lynn

  Create, train, and evaluate alternative models.

  Determine searching mechanisms.

- Writing the repeatable ETL report

  ETL report composed by S. Wainright as datasets become clean and available. All datasets should be cleaned by 5 Aug.

  Links to dataset and various files stored in SharePoint.

- Writing the presentation

  The PowerPoint layout by J. Lynn. Initial set up beginning on 3 Aug. finalized by 17 Aug.

17 – 18 Aug

- Writing the technical report

  Writing technical report begins on week two of the project via rough draft file and completed by 18 Aug.

- Coming up with a plan for presenting the presentation

  Dry run and peer presentation review completed by 18 Aug.

19 Aug

- Presentation
- Double check all deliverables are properly submitted.

Deep learning by J. Lynn

Machine learning by S. Wainright

Streaming by T. Fekadu

Demographic insight by H.Muscha

URL of planner:

https://tasks.office.com/genesis10.com/Home/PlanViews/oJYdO5Te_0SSzogH13WVxWQABdjU?Type=PlanLink&Channel=Link&CreatedTime=637951572100190000

Project flow diagram created by J.Lynn; providing an outline for the project management plan and plan maintained in Microsoft Planner.