|  |  |
| --- | --- |
| **Project Track** | Data Engineering |
| **Domain Track** | Healthcare |
| **Project Name** | Colorectal Cancer |
| **Project Team Members & Roles** | (These are just ideas)  Project Manager: Bobbi F.  Git Owner: Dawn R.  Data Collector: Linda J., Bobbi F. Dawn. R.  Database Owner: Linda J.  Slide Owner: Dawn. R, Linda J., Bobbi F.  …. |
| **Communication Methods & Out of Class Meeting Times** | |
| **Primary:** Team Slack Channel & Zoom  **Emergencies & Escalations:** Slack  **Out of Class Meeting Times:**   * TBD | |
| **Project Description & Motivation - Why is this a problem worth exploring?** | |
| Colorectal Cancer is the second leading cause of Cancer Deaths in the U.S. approximately 150,00 people will be diagnosed this year. | |
| **Top 3 Datasets Being Pursued** | |
| * Colorectal – cancer patients. <https://www.kaggle.com/datasets/amandam1/colorectal-cancer-patients?resource=download> * Colorectal – deidentified colorectal data.   <https://figshare.com/articles/dataset/supplementary_data-_deidentified_colorectal_data_xlsx/23118239>   * Dataset 3   + TBD | |
| **Top 3 Questions to Answer via Visualizations (Data Viz Track Only)** | |
| * Is there a correlation between tumor size, age, sex? * Is there a correlation between age and cancer stage? * What is the degree of aggressiveness based on age, sex, | |
| **Top 3 Javascript or Python Libraries Being Pursued** | |
| * Python: Pandas, Matplolib, Numply   + Sorting, Maps and * Matplolib, * Maps   + mathematical operations * New library TBD | |
| **Top Database Being Pursued** | |
| * Database Postgres   + Knowledge in Postgres | |