|  |  |  |  |
| --- | --- | --- | --- |
| Project Brief | | | |
| Title | Project 5: Medium [DS+, DE+++] My first Data Lake: from Data Ingestion to Data Visualization, Build a Data Lake with Insights from A to Z | Date | August 19,2020 |
| Introduction/Overview | Faced with the COVID-19 pandemic and the massive challenge it entails for global health and economies, it is imperative that social sciences be mobilized to measure, analyze and used to explain ordinary people’s perceptions of and behavioral reactions to the coronavirus (SARS-CoV-2) disease. It is equally essential that attitudes towards policies on several issues – health, the economy, civil liberties – and towards governments and institutions be investigated and documented.  The project will investigate representations, attitudes and reactions among the general public in the context of the COVID-19 pandemic in several countries exposed to the coronavirus. Contrary to conventional research projects, this proposal will be implemented over a short period of time and offer relevant and immediate deliverables for academics and public decision-makers. | | |
| Project Objectives | 1. This project aims to develop content in the COVID 19 category and the weather (collect in real-time) as a response to the COVID-19 pandemic. To provide data (in real time) on the dynamics of perceptions and (social) representations among different groups of people, based on stratified representative samples. 2. We intend to populate the site with practical, credible, and thought-provoking information on all aspects of doing a correlation between the diagnosis of COVID-19 and weather. 3. To provide insight as to why people are likely to comply with public recommendations under certain circumstances. 4. To provide data study the conditions under which the post-COVID-19 period is likely both to increase fragmentation within societies and to affect social cohesion. | | |
| Project Scope | In Scope   * Provide data from multiple data sources * Provide a conclusive report. * We will have different aspects of predictions that could help us to find the Correlation in different ways. To achieve some analysis, we will be dividing the data based on the area in France. All the predictions/analyses here are with respect to France region only.  We will use MeteoNet (An open weather Dataset by METEO FRANCE) which will act as real-time and COVID-19 from Kaggle (collect in batch). * We will deploy a lambda architecture (1 batch layer + 1 stream layer) for ingesting some data. * Storing the data in an HBase database as an example. * Plugging a Tableau Software on this data to do some great data visualization.   Out of Scope   * Provide detailed analysis of the data | | |
| Final Product | By the final deadline (listed below) we will have achieved the results as mentioned below:   * Identify the highest affected areas in France due to COVID-19. * Identify the coldest region and hottest region in France. * Compute the number of COVID cases in each region of France (based on available data in the dataset). * Correlate the relationship between COVID-19 and Weather conditions in France * Compute the death percentage in each region. | | |
| Success Factors | * Ability to generate data/information that can be easily analyzed. * Create sample analyses that are meaningful and easy to understand. * Completion of the project within given timelines. | | |
| Timeline |  | | |
| |  |  | | --- | --- | | **Title** | **Date** | | Project Brief | Completed by Friday, August 21 | | Professor review on project brief (acceptance/decline) | Completed by Friday, August 28 | | Identifying/assigning the roles of each team member | Completed by Friday, September 04 | | Data ingesting, Treating data and redefining data | Completed by Friday, September 11 | | Data pipeline and process the data for data visualization | Completed by Friday, September 18 | | Data Visualization using R or Tableau | Completed by Friday, September 25 | | Conclusion of project/ Final Submission | Completed by Monday, September 28 | | | | |