Duke Datathon 2020: The Economic Impact of COVID-19

Problem Statement: The COVID-19 pandemic has changed virtually every facet of our lives. As the months continue to pass by, we are finally beginning to witness and understand the devastating long-term economic implications of this pandemic and the various shutdowns that it has necessitated. In this Datathon, teams will be tasked with using a data-driven approach to answer the following question: which cities around the world will feel the economic impact of COVID-19 most acutely over the next 1,2, and 5 years? It is up to you to choose how many cities to present in your analysis. We encourage you to consider many different factors while developing this response, but here are some suggestions (it is not required that you explicitly consider these factors):

- Which cities are experiencing the largest changes in population due to the pandemic? Could this be due to some cities' proximity to desirable suburban or rural areas?
- Are there different shifts in various parts of the same city due to class divisions?
- Which cities depend on economic activities that heavily rely on in-person interaction and cannot be transitioned effectively to a remote setting?
- Which cities rely on industries that are facing long-term threats due to the coronavirus pandemic?
- Which cities may see the greatest economic growth due to COVID? What are the driving factors behind such growth?

To get you started, here are a few links to openly accessible datasets that may be of use. You are permitted to use other free, openly accessible data sources in your analyses as well, but please make sure to cite them in your report.

- COVID-19 Community Resilience Data (Click on the *Download Data Tables* link at the bottom of the page):
    - https://www.census.gov/data/experimental-data-products/community-resilience-estimates.html
- COVID-19 Case Counts and Death Counts (Click on *Export* at the top right of the page):
    - https://data.cdc.gov/NCHS/Provisional-COVID-19-Death-Counts-by-Sex-Age-and-S/9bhg-hcku
- Google Movement Data:
    - https://www.google.com/covid19/mobility/
- Unemployment Claims Data:
    - https://oui.doleta.gov/unemploy/claims.asp
- Large variety of COVID-19 Data (note: these are more open-ended datasets that contain a very wide variety of data):
    - https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge
    - https://www.kaggle.com/roche-data-science-coalition/uncover

Report Guidelines:

1. Please assign one person in your group to be in charge of your submissions. Please make sure to keep track of your submission number in CMT.
2. You must submit a report consisting of an abstract, introduction (including literature review), methodology, analysis, findings, and conclusion.
3. Your report may not be longer than 4 pages in length. You may have an optional appendix (maximum 2 pages), however, reviewers are not required to read this section.
4. You may have 2 pages of references that will not count against your submission.
5. You must submit your code (in a .zip file) and it must be runnable in order to be considered for a submission.
6. Submissions may be in LaTeX or Word, but must be submitted via a .pdf document.

Datathon Schedule

Zoom meeting (for all workshops): https://duke.zoom.us/j/92161233358

Oct 31 9am Kick-off Event (problem statement password released!)

9:30am Dataiku: Using DSS Software to win a Datathon

Our philosophy at Dataiku is that cross-collaboration leads to more successful data projects, and we seek to build a common understanding of the concepts and tools used to turn data into actionable business insights. Dataiku is proud to present this workshop for students & faculty to learn more about Dataiku Data Science Studio (DSS). We will discuss how data science concepts can be applied to real world problems, and help drive impactful change in today's society.

11am HackDuke: How to Start and Collaborate on a Datathon Project

Whether it's your first time participating in a datathon or your fiftieth, virtual events are new to all of us. Join HackDuke to learn about how to effectively work with your teams virtually! We'll give you tips and tricks on how to coordinate different time zones, use Tools like Git and Notion to stay organized, and talk about best practices for virtual collaboration.

12pm Duke Applied Machine Learning: Machine Learning In Practice

The Duke Applied Machine Learning Group is an organization that builds out technical solutions for startups, larger corporations, the department of defense, and universities. Over the last year, through their core organization and the Phoenix Project, they have completed nearly 100 projects in Machine Learning, Data Science, and Software. They are going to talk with us about the development process in building Machine Learning solutions in practice, the process of working with clients, and understanding the difficulties in complex data science / ML problems.

1pm Zhuoqun Wang: Intro to Dimensionality Reduction and Variable Selection

Dimensionality reduction and variable selection are two closely related topics that are frequently encountered when analyzing multivariate data, especially when we have highly correlated features and when we want to understand the contribution of the features on the outcome. I will first introduce some classic tools for dimensionality reduction for exploratory data analysis, including PCA and NMDS, where NMDS is particularly useful for visualizing data that does not lie in R^p. Then I will focus on the regression setting and go through some basics of: Lasso (for variable selection), latent factors (for capturing the unobserved variability in data / explaining high dimensional data with low dimensional structure) and Gibbs sampling (for inference under lasso or latent factor model).

All Day Mentor Help and Guidance

Go to https://duke.zoom.us/j/97664784940 and ask your question! Alternately, you can also send your question into the #mentor-help channel to crowdsource a response! Join the Slack here.

Nov 1 9am Submissions Due at dukeml.org/submit

You will work in teams of up to four members. You must submit your submission (and only one submission per group) to https://dukeml.org/submit by 9am on Sunday, November 1st. Multiple submissions will be disqualified.

1. Please assign one person in your group to be in charge of your submissions. Please make sure to keep track of your submission number in CMT.
2. You must submit a report consisting of an abstract, introduction (including literature review), methodology, analysis, findings, and conclusion.
3. Your report may not be longer than 4 pages in length. You may have an optional appendix (maximum 2 pages), however, reviewers are not required to read this section.
4. You may have 2 pages of references that will not count against your submission.
5. You must submit your code (in a .zip file) and it must be runnable in order to be considered for a submission.
6. Submissions may be in LaTeX or Word, but must be submitted via a .pdf document.

Mentor Help

To get help and guidance from mentors, go to **https://duke.zoom.us/j/97664784940** and ask your question! You can be placed into a breakout room with a mentor to get 1:1 help. Alternately, you can also send your question into the **#mentor-help** channel to crowdsource a response! Join the Slack here.