Clustering Analysis of COVID-19 Responses by County to inform Future Pandemic Prevention

The Eastcoast Eagles: Andrew Faris, Jyothi Karra, Morteza Maleki, Ruchi Patel, Betsy Thorne, Tony Bakshi

February 2022

Context

COVID-19 has impacted people, hospitals, communities, families, and much more.

Given the available data, there is an opportunity to leverage modern machine learning to identify improvement opportunities for each county in the USA.

78.4M US Cases*

935K US Deaths**

*https://www.nytimes.com/interactive/2020/us/coronavirus-us-cases.html
**https://ourworldindata.org/coronavirus-data

Objectives of the study

- Identify "high-performing" groups of counties
- Execute summary statistics about why the groups might be “high-performing”, and analyze “lower-performing” groups of counties to suggest improvement opportunities
- Use data-driven insights to instill new practices into how to manage COVID-19

How is COVID control / analysis done today?

- Conservative / broad sweeping mandates to prevent transmission
- Big emphasis on cases, deaths, and vaccine doses
- Descriptive and diagnostic analytics, instead of predictive and prescriptive analytics

Our approach

- Ingest large spectrum of data (COVID statistics, demographic, behavioral data) at county level
- Feature engineer and normalize data
- K-Means clustering algorithm
- Visualization of K-Means clusters utilizing Principal Component Analysis (top two principal components)
- Analyze clusters using summary statistics
- Identify recommendations for each cluster to help with COVID control

What is innovative and new?

- K-Means utilizing a broad integrated data set at county level (i.e., more granular)
- Focus on predictive / prescriptive actions for counties to take

Who cares? What difference will this make?

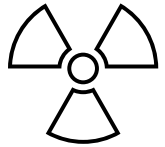
Who Cares?

- Public officials and policymakers
- Everyday US citizens
- Pandemic researchers / scientists

What difference will this make?

- Potentially can save future lives
- Reduce strain on hospital / health care systems
- Mitigate negative economic impact

Risks and payoffs



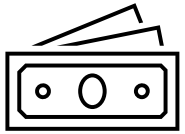
Risks

- Misuse of demographic data / bias
- Unreasonable recommendations

Payoffs

- Data-driven approach to improve human lives
- Reduce economic impact on society
- Reduce strain on healthcare system

Costs and effort estimates



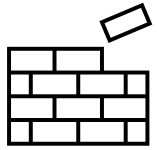
Cost

- Computing power
- Time (est. $20,000)

Effort Estimates

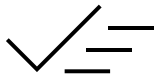
- Duration: 8 weeks
- Hours: 6 resources at 20 hours per week

Monitoring progress and success



Midterm check

- Data is available and can be integrated
- Logical groupings / clusters established
- KPIs: # of reasonable clusters, # of available features



Final check

- Reasonable summary statistics by cluster
- Reasonable / actionable recommendations by cluster to be transcended to each county
- KPIs: # of recommendations by cluster, # of next steps for study

Effort Estimates*/Next Steps

Name	Task(s)	Start Date	Target date
Andrew Faris	Model Dev, Report Writing, Poster Building	1/25/2022	4/23/2022
Jyothi Karra	Data Engineering, Report Writing, Poster Building	1/25/2022	4/23/2022
Morteza Maleki	Data Engineering, Report Writing, Poster Building	1/25/2022	4/23/2022
Ruchi Patel	Model Dev, Report Writing, Poster Building	1/25/2022	4/23/2022
Betsy Thorne	Visualization, Report Writing, Poster Building	1/25/2022	4/23/2022
Tony Bakshi	Model Dev, Report Writing, Poster Building	1/25/2022	4/23/2022

*All team members have contributed a similar amount of effort.

Bibliography

1. David Holtz, Michael Zhao, Seth Benzell, Cathy Cao, Mohammad Rahimian, Jeremy Yang, Jennifer Allen, Avinash Collis, Alex Moehring, Tara Sowrirajan, Dipayan Ghosh, Yunhao Zhang, Paramveer S. Dhillon, Christos Nicolaides, Dean Eckles, Sinan Aral, Proceedings of the National Academy of Sciences Aug 2020, 117 (33) 19837-19843; DOI: 10.1073/pnas.2009522117
2. Painter, Marcus and Qiu, Tian, Political Beliefs affect Compliance with Government Mandates (March 8, 2021). Journal of Economic Behavior and Organization, Forthcoming, http://dx.doi.org/10.2139/ssrn.3569098
3. The relationship between vaccination rates and COVID-19 cases and deaths in the USA https://systems.jhu.edu/research/public-health/covid-19-vaccine/?fbclid=IwAR21qH7AMfCjDS-PE1BTJyQO4vodMUZhWAdkmHp5v653hDSgZFTqusrKAEY Authors: Ensheng Dong and Lauren Gardner
4. Viswanath, K., Bekalu, M., Dhawan, D. et al. Individual and social determinants of COVID-19 vaccine uptake. *BMC Public Health* **21,** 818 (2021). https://doi.org/10.1186/s12889-021-10862-1
5. Abedi, V., Olulana, O., Avula, V. *et al.* Racial, Economic, and Health Inequality and COVID-19 Infection in the United States. *J. Racial and Ethnic Health Disparities* **8,** 732–742 (2021). https://doi.org/10.1007/s40615-020-00833-4
6. Ran Liu, Gabriel Miao Li, Hesitancy in the time of coronavirus: Temporal, spatial, and sociodemographic variations in COVID-19 vaccine hesitancy, SSM - Population Health, Volume 15, 2021, 100896, ISSN 2352-8273 https://doi.org/10.1016/j.ssmph.2021.100896
7. Kelly, B.J., Southwell, B.G., McCormack, L.A. *et al.* Predictors of willingness to get a COVID-19 vaccine in the U.S. *BMC Infect Dis* **21,** 338 (2021). https://doi.org/10.1186/s12879-021-06023-9
8. Covid-19 by Race and Ethnicity: A National Cohort Study of 6 Million United States Veterans Christopher T. Rentsch, Farah Kidwai-Khan, Janet P. Tate, Lesley S. Park, Joseph T. King Jr., Melissa Skanderson, Ronald G. Hauser, Anna Schultze, Christopher I. Jarvis, Mark Holodniy, Vincent Lo Re III, Kathleen M. Akgün, Kristina Crothers, Tamar H. Taddei, Matthew S. Freiberg, Amy C. Justice medRxiv 2020.05.12.20099135; doi: https://doi.org/10.1101/2020.05.12.20099135
9. Covid-19 by Race and Ethnicity: A National Cohort Study of 6 Million United States Veterans , Christopher T. Rentsch, Farah Kidwai-Khan, Janet P. Tate, Lesley S. Park, Joseph T. King Jr., Melissa Skanderson, Ronald G. Hauser, Anna Schultze, Christopher I. Jarvis, Mark Holodniy, Vincent Lo Re III, Kathleen M. Akgün, Kristina Crothers, Tamar H. Taddei, Matthew S. Freiberg, Amy C. Justice medRxiv 2020.05.12.20099135; doi: https://doi.org/10.1101/2020.05.12.20099135
10. Chen, J., Vullikanti, A., Santos, J. *et al.* Epidemiological and economic impact of COVID-19 in the US. *Sci Rep* **11,** 20451 (2021). https://doi.org/10.1038/s41598-021-99712-z
11. Chen, J., Vullikanti, A., Santos, J. *et al.* Epidemiological and economic impact of COVID-19 in the US. *Sci Rep* **11,** 20451 (2021). https://doi.org/10.1038/s41598-021-99712-z
12. Shashank Reddy Vadyala, Sai Nethra Betgeri, Eric A. Sherer, Amod Amritphale, Prediction of the number of COVID-19 confirmed cases based on K-means-LSTM, Array,Volume 11,2021,100085,ISSN 2590-0056,https://doi.org/10.1016/j.array.2021.100085. (https://www.sciencedirect.com/science/article/pii/S2590005621000333)
13. .Juniar Hutagalung and Ni Luh Wiwik Sri Rahayu Ginantra and Gita Widi Bhawika and Wayan Gede Suka Parwita and Anjar Wanto and Pawer Darasa Panjaitan, title = {{COVID}-19 Cases and Deaths in Southeast Asia Clustering using K-Means Algorithm, doi = {10.1088/1742-6596/1783/1/012027},url = {https://doi.org/10.1088/1742-6596/1783/1/012027}
14. Daniel J. McDonald, Jacob Bien, Alden Green, Addison J. Hu, Nat DeFries, Sangwon Hyun, Natalia L. Oliveira, James Sharpnack, Jingjing Tang, Robert Tibshirani, Valérie Ventura, Larry Wasserman, Ryan J. Tibshirani, Proceedings of the National Academy of Sciences Dec 2021, 118 (51) e2111453118; DOI: 10.1073/pnas.2111453118
15. Zhao H, Merchant NN, McNulty A, Radcliff TA, Cote MJ, Fischer RSB, et al. (2021) COVID-19: Short term prediction model using daily incidence data. PLoS ONE 16(4): e0250110. https://doi.org/10.1371/journal.pone.0250110
16. Wanberg, C. R., Csillag, B., Douglass, R. P., Zhou, L., & Pollard, M. S. (2020). Socioeconomic status and well-being during COVID-19: A resource-based examination. Journal of Applied Psychology, 105(12), 1382–1396. https://doi.apa.org/fulltext/2020-77456-001.html
17. Hippisley-Cox J, Coupland C A, Mehta N, Keogh R H, Diaz-Ordaz K, Khunti K et al. Risk prediction of covid-19 related death and hospital admission in adults after covid-19 vaccination: national prospective cohort study *BMJ* 2021; 374 :n2244 doi:10.1136/bmj.n2244
18. Falato, Antonio and Goldstein, Itay and Hortaçsu, Ali, title = "Financial Fragility in the COVID-19 Crisis: The Case of Investment Funds in Corporate Bond Markets", doi = {10.3386/w27559}, URL = "http://www.nber.org/papers/w27559",
19. Miller, Amalia R and Segal, Carmit and Spencer, Melissa K, Effects of COVID-19 Shutdowns on Domestic Violence in US Cities, doi = {10.3386/w29429},URL = "http://www.nber.org/papers/w29429",
20. Chang, Tom and Jacobson, Mireille and Shah, Manisha and Pramanik, Rajiv and Shah, Samir B, Financial Incentives and Other Nudges Do Not Increase COVID-19 Vaccinations among the Vaccine Hesitant, doi=10.3386/w29403,URL http://www.nber.org/papers/w29403
21. Md. Zubair, MD.Asif Iqbal, Avijeet Shil, Enamul Haque, Mohammed Moshiul Hoque, Iqbal H. Sarker, An Efficient K-means Clustering Algorithm for Analysing COVID-19, 2020, https://doi.org/10.48550/arXiv.2101.03140