On the ranking of viable American states for setting up a new restaurant chain for "Dadadon"

Y. S. Teo*

We propose and explain an optimal selection of states out of all 50 (and the capital) in the United States that are suitable for opening a new chain of "Dadadon", a Michelin 3-star Spanish restaurant. This optimal choice is based on several relevant societal factors, namely the quality of life, high-tech labor workforce, crime rate, the support for wholesaler industries, and recent financial news reports.

I. INTRODUCTION

Alejandro Banderas is a famous celebrity masterchef who appears regularly on both Spanish and American television shows and owns three successful chains of restaurants "Dadadon" in New York City. The Spanish gourmet house "Dadadon" belongs to a class of moderately to highly luxurious restaurants that is popular amongst both working-class professionals as well as the general affluent public.

The success of the 3-star chains primarily hinges on Alejandro's highly-trained and creative culinary intuitions and his demand for choosing only the finest ingredients for all his signature dishes. One of the most highly rated dishes, Paella de Maricos, is cooked with fresh chicken, live seafood and local field beans.

Upon the request of Alejandro to set up a new chain of "Dadadon" in another American state similar to New York, we shall proceed to incorporate all key societal factors that contribute to its overall business. Situated in the state New York, these are its quality of life, healthy population of high-tech professional work force, a well-connected transportation system and manageable crime rates. At the behest of Alejandro to reduce the cost of express deliveries of fresh produce, we shall also provide a list of appropriate American states that additionally support thriving local wholesaler industries of various crucial raw items.

The report is organized as follows. In Sec. II, we shall describe, in detail, all relevant data that measures the aforementioned four important criteria to shortlist and rank an optimal set of American states, which are collected from the Wikipedia, Bureau of Labor Statistics (BLS), Federal Bureau of Investigation (FBI) and a paper published in Ecological Economics. Next, in Sec. III, we explain the technique of K-means clustering employed to cluster all 50 American states (and District of Columbia) according the quality of life, high-tech work force and crime rates, the outcome of which is then later sorted according to the status of wholesaler industries. The results of these methods shall be presented in Sec. IV and discussed in Sec. V. Finally, Sec. VI concludes the report.

II. DATA

We begin by describing the necessary datasets to measure (A) quality of life, (B) high-tech professional work force and transportation efficiency, (C) crime rates, and (D) wholesaler market strength. These four social factors are taken to be primary in influencing the business of "Dadadon". The assumptions of measures for these social factors and the different sources from which all corresponding datasets are gathered are stated in the subsections below.

A. Quality of life

The quality of life is measured with two popular indicators—the Human Development Index (HDI) [1] and Genuine Progress Indicator (GPI) [2], both of which perform more homogenously in indicating societal progress and well-being than the conventional GDP measure [4]. The table of HDI values for every American state for the year 2017 is extracted from [3]. The GPI values, on the other hand, are inferred from [2].

B. High-tech work force

In this report, a high-tech work force and efficient transportation consists of employments in the information, service, finance, professional and business, education and health, leisure and hospitality as well as transportation and warehousing industries.

To quantitatively measure the intensity of the workforce, we look at the average annual establishments, average employments and total annual wages for a given year in each American state. All latest relevant data for the year 2017, including a table of Federal Information Processing Standards (FIPS) area codes, are extracted from [5].

C. Crime rates

The crime-rate dataset includes a comprehensive list of different types of recorded crime-rates and population of each American state. The latest complete dataset for the year 2014 is obtained from the uniform crime reporting online research tool in [6].

D. Wholesaler market strength

Since "Dadadon" heavily relies on fresh ingredients, seafood, meat and vegetable raw products, in order to

* ys_teo@snu.ac.kr

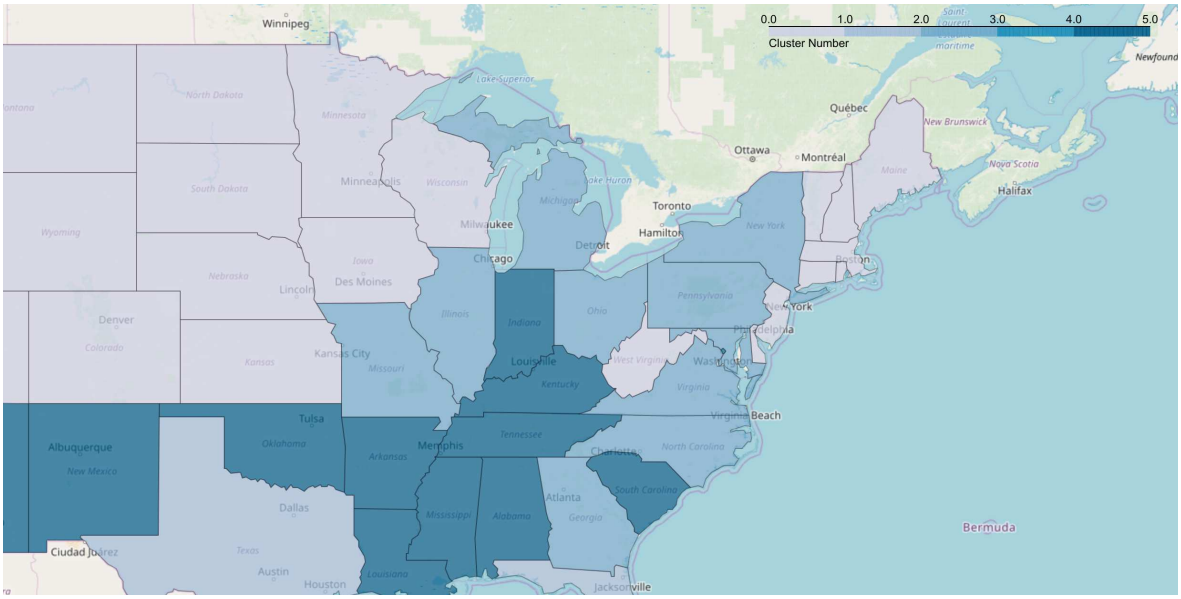

FIG. 1. A choropleth map of the United States, with focus on the relevant cluster number 2 that includes the state of New York.

sufficiently reduce the cost of intercity express delivery and ingredient fees, we propose to rank the viable American states according to the strength of local wholesaler industries. These data are again available in [5], where market strength is measured by the grand total annual wages per average employment or capita for a fair comparison.

E. Data processing

All datasets from Secs. II A to II C are organized and merged to a large dataset that represents the important and relevant social aspects of all 50 American states and the capital D.C that could affect the business of "Dadadon". The dataset from Secs. II D is separately trimmed and manipulated to contain the state names and grand annual total wages per capita.

III. METHODOLOGY

We proceed in two steps to analyze the processed data. In the first step, to the large dataset that contains social-aspect measurements of every American state, we use the machine-learning technique called the K-means clustering method [7, 8] to group all states in accordance with their similarities in terms of these measurements after a standard normalization and transformation. This algorithm requires a choice of k, which is the number of clusters to be fixed for the algorithm.

In order to find the optimal k, we investigate the *inertia* of the clusters, which is the sum of distances between different pairs of clusters. As the inertia always decreases with increasing k, a useful optimal k that is not too large shall be defined as one beyond which there is an abrupt change in the inertia curvature in k.

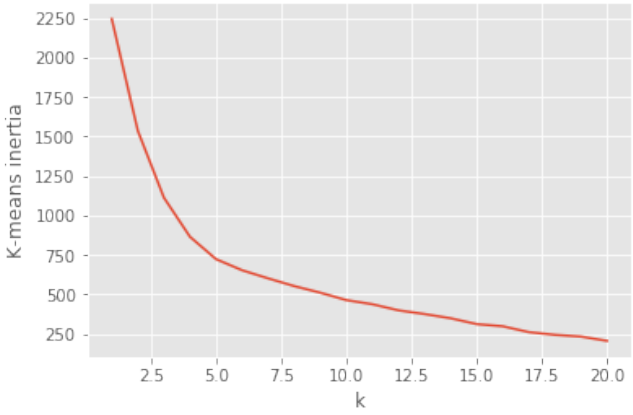

FIG. 2. A plot of K-means inertia against the number of clusters k. A kink in the curvature occurs at $k = 5$, which is taken to be the optimal value for our analysis.

In the second step, we perform data sorting on the dataset measuring wholesaler strength. All states are sorted in decreasing order of the grand total annual wages per capita. This sorted set is then intersected with the computed state-cluster results to obtain the optimal set of American states ranked according to the wholesaler strength.

IV. RESULTS

Figure 1 shows the choropleth map of the United States that summarizes the results for the first step of our methodology. Based on K-means clustering based on societal factors described in Secs. II A through II C, the second cluster that contains the state of New York also includes nine other

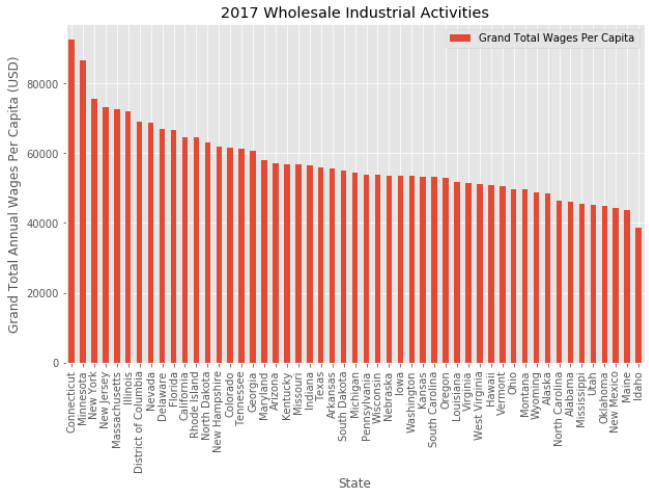

FIG. 3. A bar chart showing the grand total annual wages per capita of the relevant wholesaler activities for all states and capital in the United States.

State	Grand total wages per capita
New York	75515.38
Illinois	72086.17
Georgia	60661.81
Maryland	57987.95
Missouri	56664.26
Michigan	54452.75
Pennsylvania	53865.81
Washington	53435.88
Virginia	51512.19
Ohio	49638.18
North Carolina	46416.40

TABLE I. Sorted states in descending order of wholesaler strength for the second cluster.

American states.

A total of $k = 5$ clusters are formed using K-means clustering to group all American states and the capital. This is motivated from a simple optimization procedure showing that $k = 5$ is the useful optimal value. To this end, we compute the K-means inertia up to $k = 20$. The optimal k that is not too large can be observed in Fig. 2.

The results for the sorted wholesaler dataset is given in Fig. 3. We next take the sorted American states and capital based on this dataset and rank the ten states in the second cluster in descending order of wholesaler market strength. The final set of viable American states is tabulated in Tab. I.

V. DISCUSSION

Our analysis shows that none of the states in cluster 2 possess wholesaler market strengths stronger than New York. The best option would thus be to select the state with the next strongest wholesaler market force.

We note, however, that other external factors and news reports *must* be factored in to finalize the optimal choice. A recent study in Illinois Policy [11] that reports a fiscal crisis in the Illinois state due to financial "budget gimmicks" and "misplaced spending priorities". Therefore picking Illinois would generate justifiable concerns.

In view of this recent financial concern, the next two states in line are Georgia and Maryland. We highlight that UGA Today in 2018 [9] mentioned that Georgia is projected to have a healthy economic growth in 2019, and is as such a favorable state to do business, although in [10] it was reported that Georgia falls in the 18th position in terms of fiscal health. On the other hand, in 2019, it was released in [12] that the state of Maryland comes in number 4 in the number of millionaire households, and hence the top state in the second cluster tabulated in Tab. I. A recent budget surplus report in [13] further supports this sensible alternative.

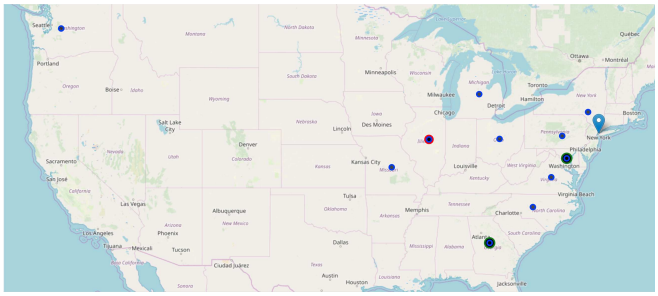

FIG. 4. A marked map showing all ten cluster-2 American states (blue circles) to set up another chain of "Dadadon", with the top three states Illinois, Georgia and Maryland supporting wholesaler industries in 2017 right after New York. The blue drop point marks New York City. Both Georgia and Maryland (circled in green) are good candidates in view of the recent poor fiscal health of Illinois (circled in red).

VI. CONCLUSION

Using machine learning and sorting methods, we analyzed the states and capital of the United States in order to suggest a list of plausible states appropriate for setting up a new chain of "Dadadon". After taking recent financial news reports into consideration, we suggest either *Georgia* or *Maryland* as the optimal state to open the new chain. The precise selection would depend on the decision by the principal shareholder, Masterchef Alejandro after further discussions. Georgia is suitable if Alejandro is confident in its reported economic projections. Otherwise Maryland is the more stable conservative choice due to its well-established fiscal status.

ACKNOWLEDGMENTS

The author is grateful to IBM and Coursera for their instructive and enriching online courses and example

exercises, and would also like to thank all peers who have helped to mark the assignments.

[1] H. Wolff, H. Chong, and M. Auffhammer, *Classification, Detection and Consequences of Data Error: Evidence from the Human Development Index*, Econ. J. **121**, 843 (2011).

[2] M.-J V. Fox and J. D. Erickson, *Genuine Economic Progress in the United States: A Fifty State Study and Comparative Assessment*, Ecol. Econ. **147**, 29 (2018).

[3] *List of U.S. states and territories by Human Development Index*, Wikipedia (2017), https://en.wikipedia.org/wiki/List_of_U.S._states_and_territories_by_Human_Development_Index.

[4] A. F. Ivković, *Limitations of the GDP as a measure of progress and well-being*, Ekonomski Vjesnik BR. 1/2016, 257 (2016).

[5] Bureau of Labor Statistics (2017), https://www.bls.gov/.

[6] *Uniform Crime Reporting (UCR) Program*, Federal Bureau of Investigation (2014), https://www.fbi.gov/services/cjis/ucr.

[7] H. Steinhaus, *Sur la division des corps matériels en parties*, Bull. Acad. Polon. Sci. **4**, 801 (1957).

[8] J. B. MacQueen, *Some Methods for classification and Analysis of Multivariate Observations*, Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability **1**, 281 (1967).

[9] UGA Today (2018), https://news.uga.edu/economic-outlook-2019/

[10] Mercatus Center (2018), https://www.mercatus.org/statefiscalrankings/georgia

[11] Illinois Policy (2018), https://www.illinoispolicy.org/reports/the-history-of-illinois-fiscal-crisis/.

[12] Baltimore Business Journal (2019), https://www.bizjournals.com/baltimore/news/2019/01/29/maryland-is-no-longer-the-top-state-for.html

[13] The Baltimore Sun (2018), https://www.baltimoresun.com/news/maryland/politics/bs-md-fiscal-surplus-20180831-story.html.