K-MEANS PROJECT

CSCI 3302 - Machine Learning & Knowledge Discovery

Joshua Kyle Aceret & Alyssa Lawton
Hawaii Pacific University

MOTIVATION OF PROJECT

- We wanted to see if tourism did have an impact on the number of COVID-19 cases in Hawaii.
  - Find any possible information in regards to COVID-19 in Hawaii, with a primary focus on trans-pacific travelers.

- Hoped to find possible trends through the use of **K-Means Clustering**.

RELATED REFERENCE WORK

- “Assessing global preparedness for the next pandemic: development and application of an Epidemic Preparedness Index”
  - K-Means Clustering was used on **2009 H1N1 influenza pandemic**.
  - Clusters were used to determine a Epidemic Preparedness Index (EPI).
  - Used to improve overall preparedness for possible future pandemics.

- “Immunophenotyping of COVID-19 and influenza highlights the role of type I interferons in development of severe COVID-19”
  - K-Means Clustering was used on blood samples from COVID-19 patients. (Mild & Severe)
  - Clusters were based on gene expression changes against a healthy donor group.

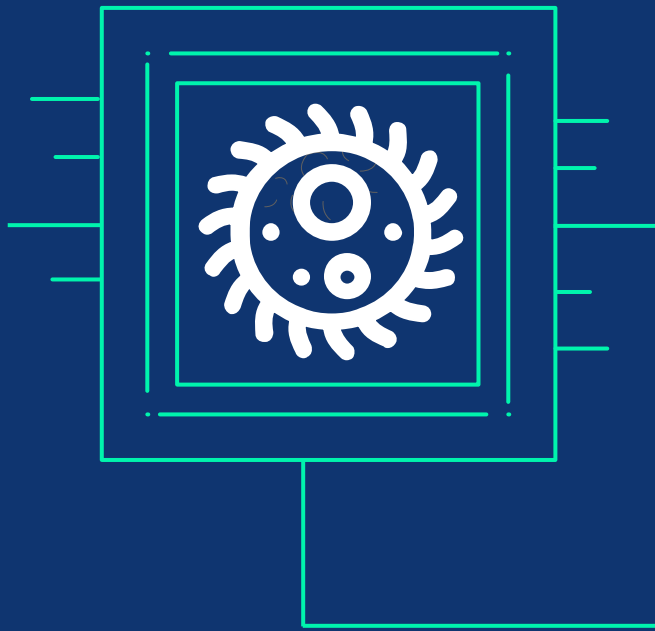
RELATED REFERENCE WORK

- “Enhancement of hepatitis virus outcome predictions with application of K-means clustering”
  - K-Means Clustering was used to determine if a patient with symptoms similar to Hepatitis B and Hepatitis C, actually has it.
  - The program had a roughly 84.85% success rate.
  - Authors note that K-Means Clustering could be used to make other medical predictions.

DATASET OVERVIEW

FEATURES	RANGE	AVERAGE
Daily COVID-19 Cases with Travel History	0 - 50 Cases	4.64 Cases
Percentage of all COVID-19 Cases with Travel History	0% - 100%	10.85%
Daily Trans-Pacific Passenger Arrivals	386 - 24,654 Arrivals	6,479.43 Arrivals

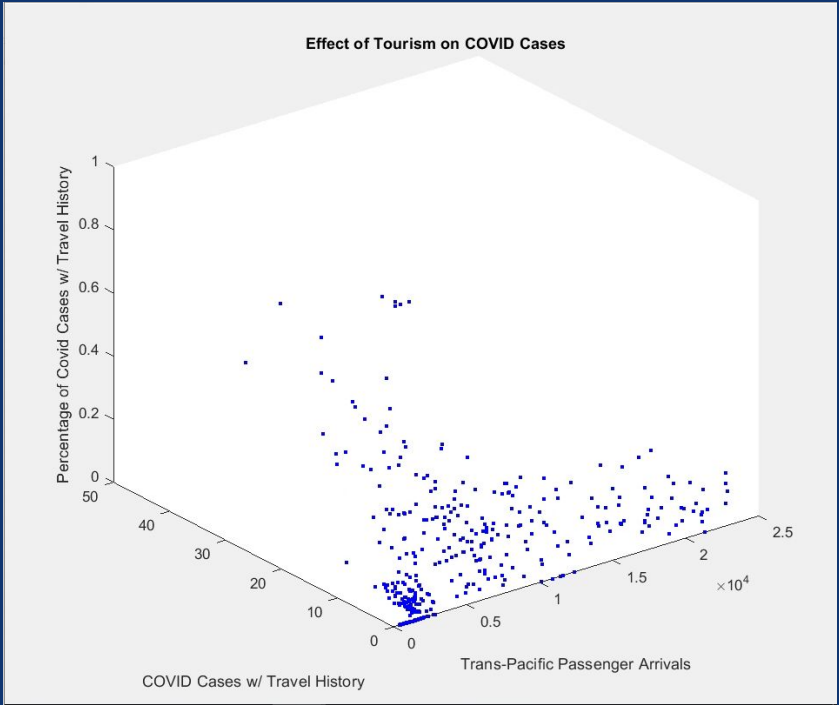
***Features - Categorized by County and Statewide*



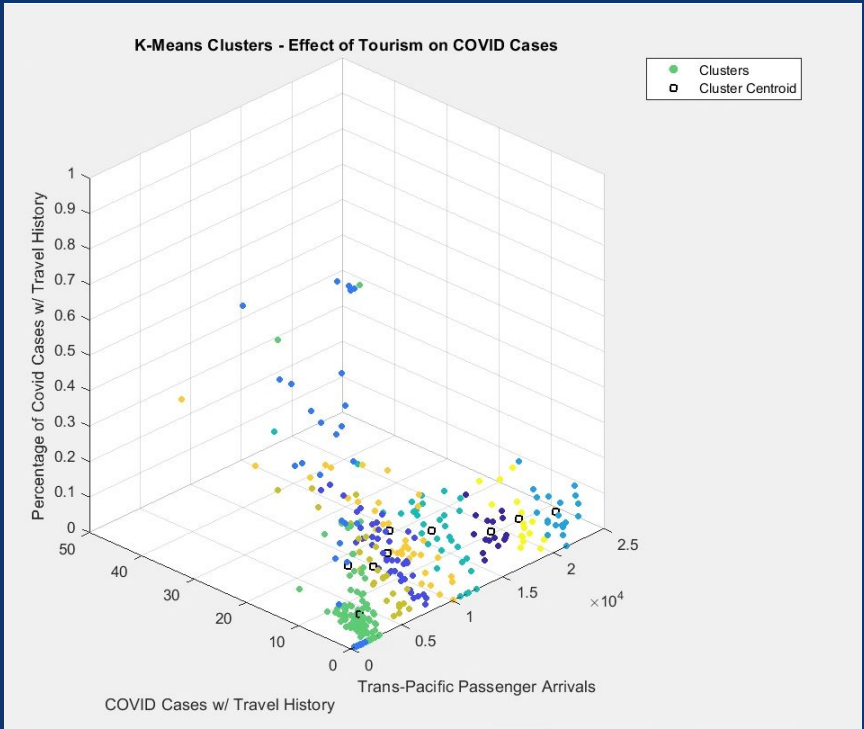
K-MEANS CLUSTERING OVERVIEW

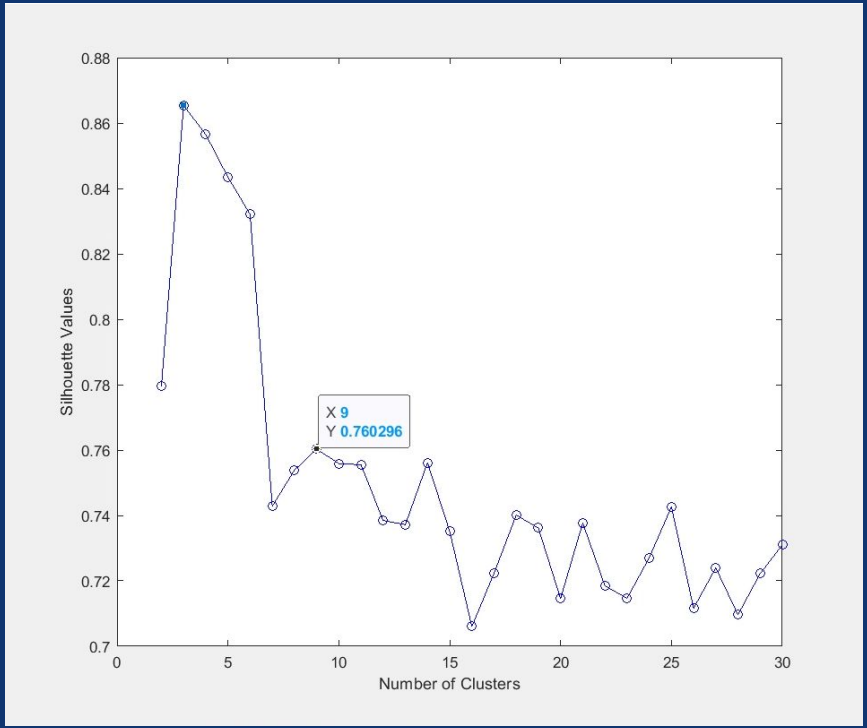
- Type of unsupervised learning, which is used on unlabeled data.
- The goal is to find groups of data with a number of groups k, based on feature similarity.
- When a new data point is added, it can be grouped with a cluster based on its features.
- Can be used to find trends in data and classify new data accordingly.

RESULTS


Effect of Tourism on COVID Cases
Percentage of Covid Cases w/ Travel History
1
0.8
0.6
0.4
0.2
0
50
40
30
20
10
0
COVID Cases w/ Travel History
0
0.5
1
1.5
2
2.5
×10^4
Trans-Pacific Passenger Arrivals

RESULTS


K-Means Clusters - Effect of Tourism on COVID Cases
Clusters
Cluster Centroid
Percentage of Covid Cases w/ Travel History
COVID Cases w/ Travel History
Trans-Pacific Passenger Arrivals
×10^4

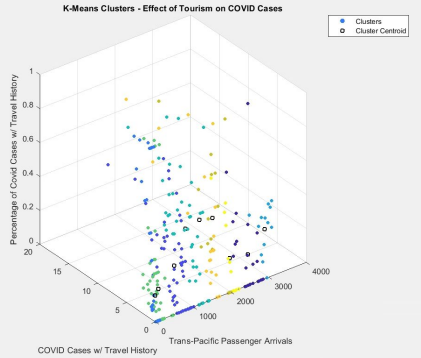

Silhouette Values
Number of Clusters
X 9
Y 0.760296

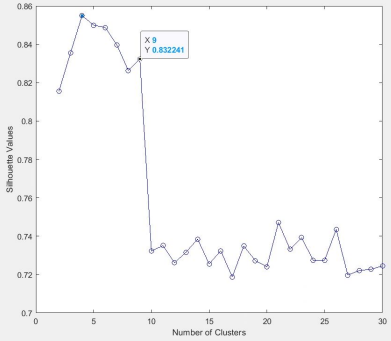
RESULTS

CLUSTER CENTROIDS	ARRIVALS	# OF TRAVEL COVID CASES	% OF TRAVEL COVID CASES
1	781	2	20.98%
2	2,050	2	5.47%
3	5,786	7	10.79%
4	7,623	8	11.73%
5	9,449	11	13.49%
6	12,065	8	11.84%
7	15,822	4	9.10%
8	18,812	4	8.36%
9	21,959	3	6.71%

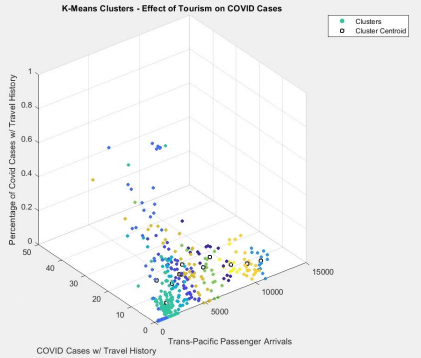
RESULTS PER COUNTY

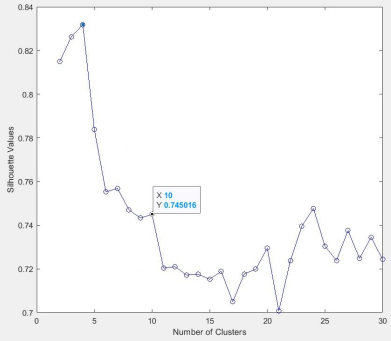
HAWAII



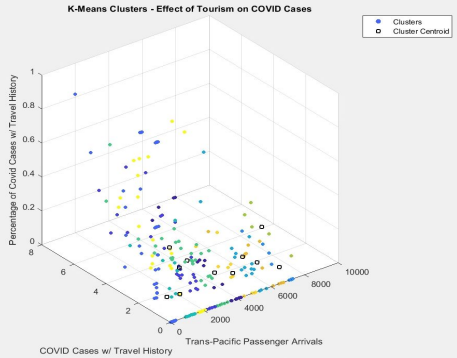


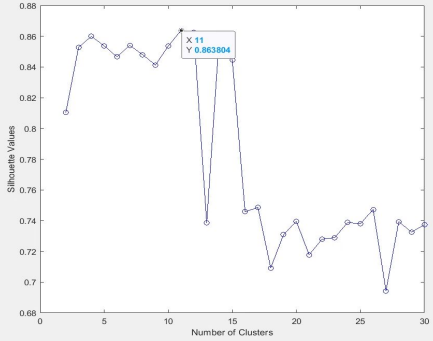
OAHU



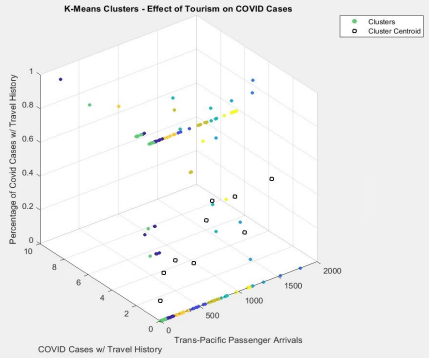


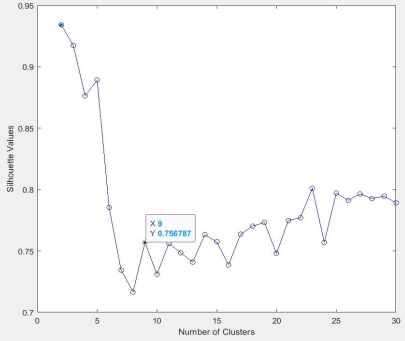
MAUI



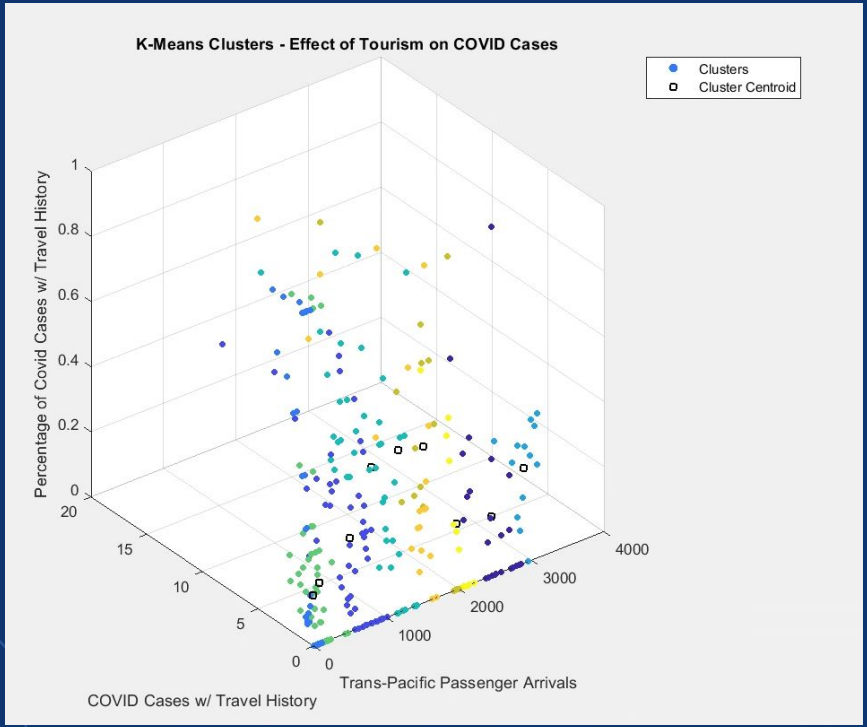


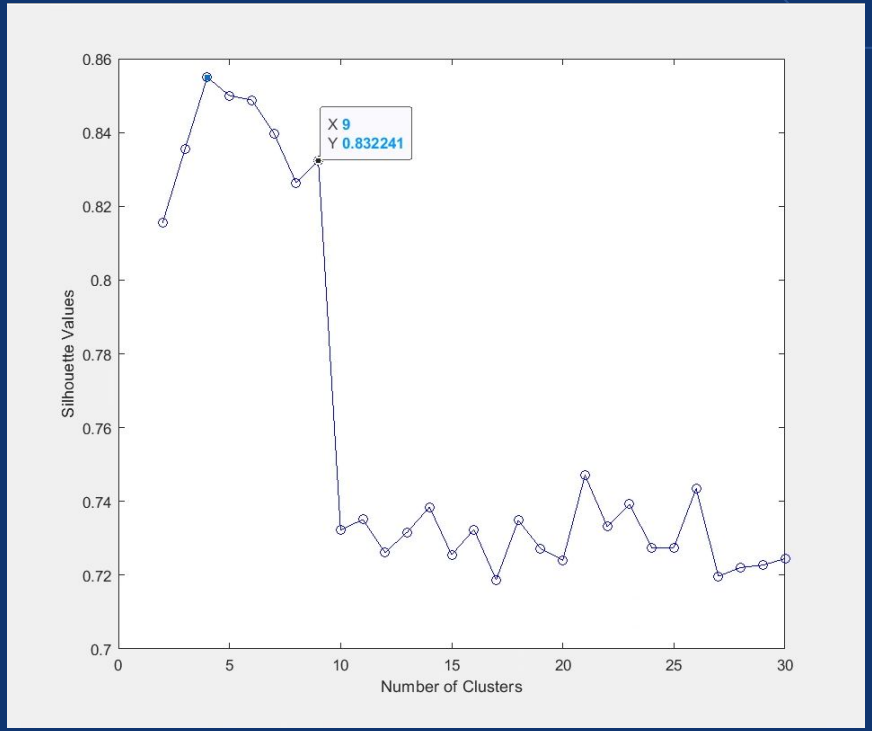
KAUAI



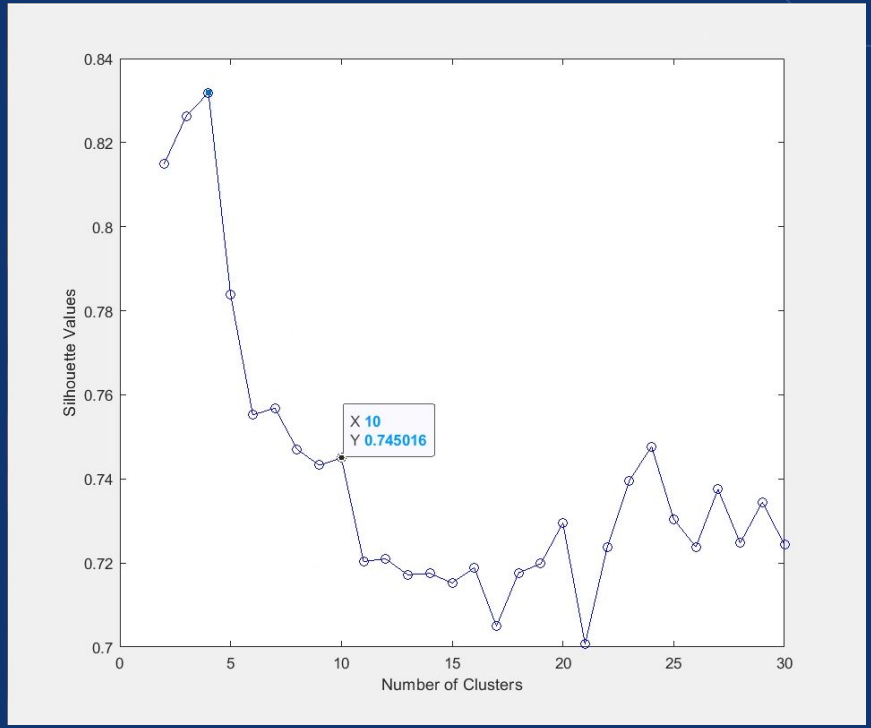


HAWAII RESULTS

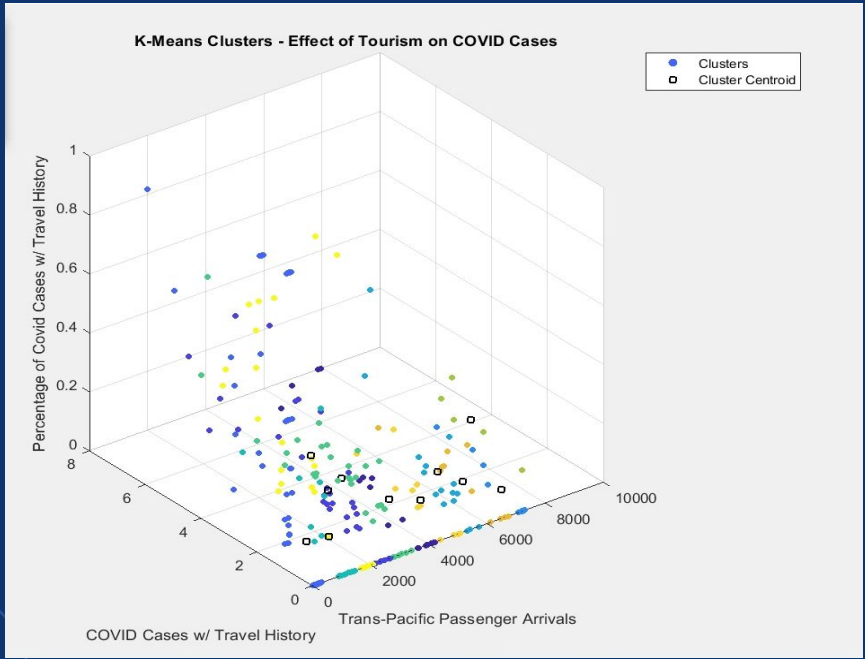


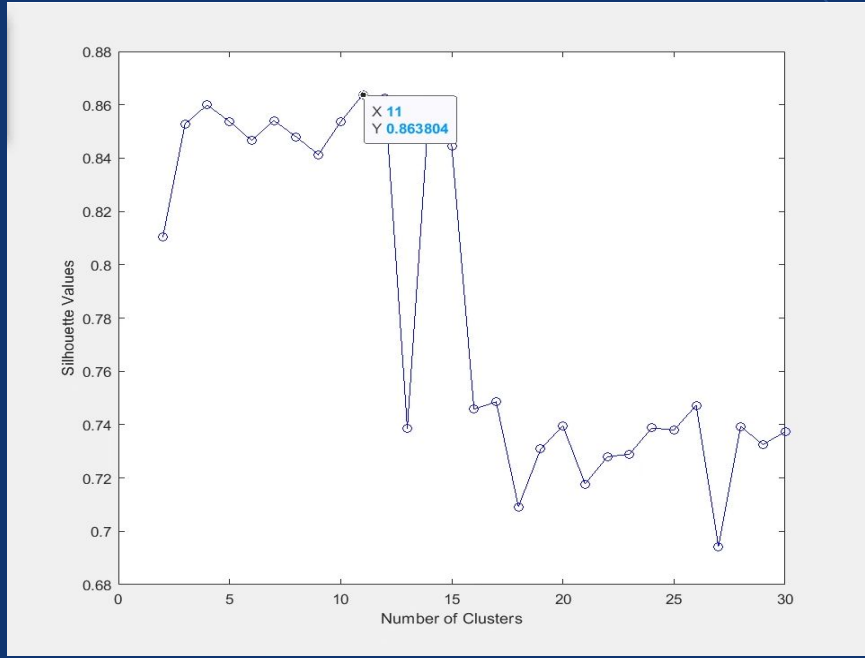


OAHU RESULTS

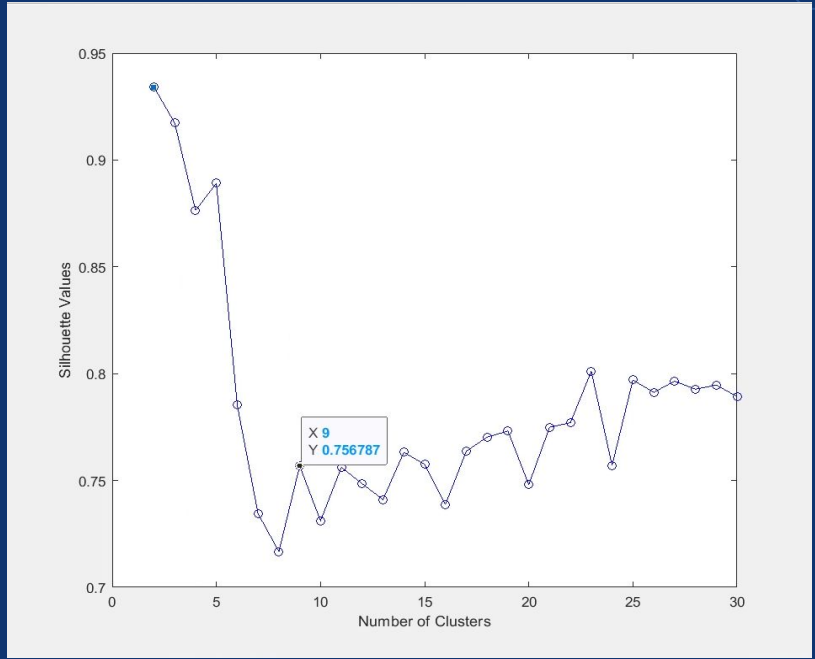


MAUI RESULTS





KAUAI RESULTS



RESULTS DISCUSSION

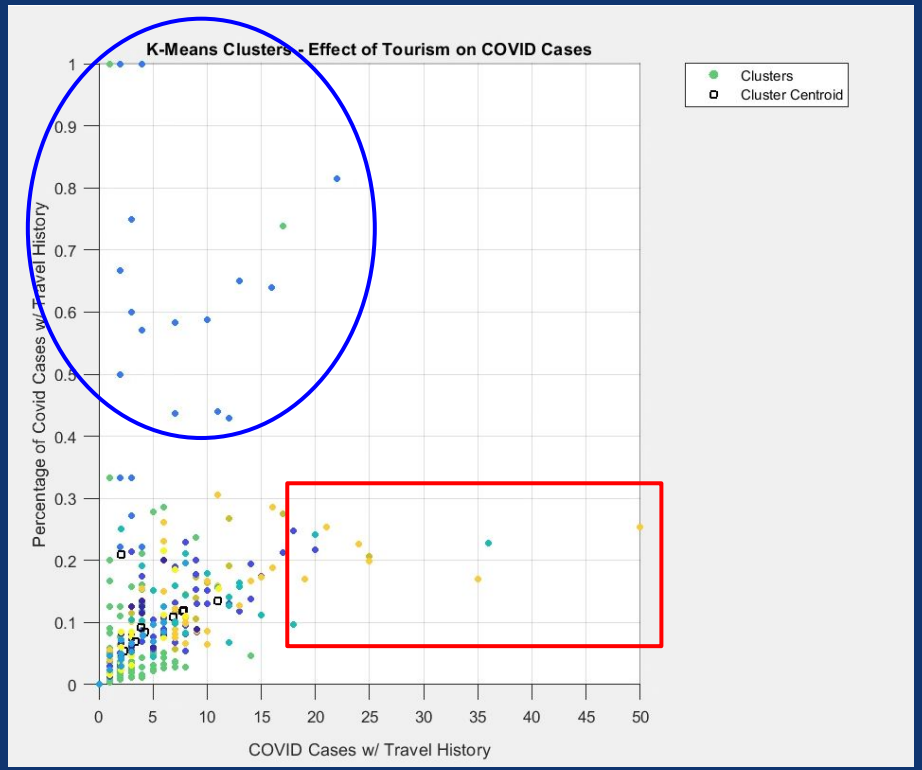
- Outliers in Red Rectangle
  - 1/5 -1/7 → Christmas
  - 1/8 - 1/12 → New Year's
- Days when most COVID-19 cases were associated with travel correspond to Major Holidays

RESULTS DISCUSSION

- Outliers in Blue Circle
  - 3/2020 - 5/2020

→ At start of COVID-19, most cases were associated with travel but few number of COVID-19 cases

- Outliers in Red Rectangle
  - 11/24 → Veterans Day
  - 12/3, 12/8 → Thanksgiving
  - 1/4 - 1/7 → Christmas
  - 1/8 - 1/14 → New Year's



CONCLUSION

- Implementation of the K-Means Clustering algorithm provided an accurate representation of the effect of tourism on COVID-19 Cases.
- Major Findings:
  - Tourism is not a major contributor to COVID-19 Cases for the overall State of Hawaii
  - On the contrary, tourism does have an impact on the island of Kauai independently

THANK YOU!

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**

RESOURCES

[1] A. Ahmad and L. Dey, “A k-mean clustering algorithm for mixed numeric and categorical data,” Data & Knowledge Engineering, vol. 63, no. 2, pp. 503–527, Nov. 2007.

[2] S. Amin, M. I. Uddin, M. A. Zeb, A. A. Alarood, M. Mahmoud, and M. H. Alkinani, “Detecting Dengue/Flu Infections Based on Tweets Using LSTM and Word Embedding,” IEEE Access, vol. 8, pp. 189054–189068, Oct. 2020.

[3] J. S. Lee, S. Park, H. W. Jeong, J. Y. Ahn, S. J. Choi, H. Lee, B. Choi, S. K. Nam, M. Sa, J.-S. Kwon, S. J. Jeong, H. K. Lee, S. H. Park, S.-H. Park, J. Y. Choi, S.-H. Kim, I. Jung, and E.-C. Shin, “Immunophenotyping of COVID-19 and influenza highlights the role of type I interferons in development of severe COVID-19,” Science Immunology, vol. 5, no. 49, pp. 1–16, Jul. 2020.

[4] B. Oppenheim, M. Gallivan, N. K. Madhav, N. Brown, V. Serhiyenko, N. D. Wolfe, and P. Ayscue, “Assessing global preparedness for the next pandemic: development and application of an Epidemic Preparedness Index,” BMJ Global Health, vol. 4, no. 1, pp. 1–9, Jan. 2019.

[5] A. M. Qahtani, B. M., H. Alhakami, S. Abuayeid, and A. Baz, “Predicting Hospitals Hygiene Rate during COVID-19 Pandemic,” International Journal of Advanced Computer Science and Applications, vol. 11, no. 12, pp. 815–823, 2020.

[6] A. Solovyov, G. Palacios, T. Briese, W. I. Lipkin, and R. Rabadan, “Cluster analysis of the origins of the new influenza A(H1N1) virus,” Eurosurveillance, vol. 14, no. 21, May 2009.

[7] M. Stricker, N. Onland-Moret, J. Boer, Y. V. D. Schouw, W. Verschuren, A. May, P. Peeters, and J. Beulens, “Dietary patterns derived from principal component- and k-means cluster analysis: Long-term association with coronary heart disease and stroke,” Nutrition, Metabolism and Cardiovascular Diseases, vol. 23, no. 3, pp. 250–256, Mar. 2013.

[8] G. Sun, Y. Hakozaki, S. Abe, N. Q. Vinh, and T. Matsui, “A novel infection screening method using a neural network and k-means clustering algorithm which can be applied for screening of unknown or unexpected infectious diseases,” Journal of Infection, vol. 65, no. 6, pp. 591–592, Dec. 2012.

[9] K. Teknomo, “K-Mean Clustering Tutorials,” K-Mean Clustering Tutorial, 2019. [Online]. Available: https://people.revoledu.com/kardi/tutorial/kMean/index.html. [Accessed: 12-Mar-2021].

[10] J. Turtle, P. Riley, M. Ben-Nun, and S. Riley, “Accurate influenza forecasts using type-specific incidence data for small geographical units,” medRxiv, Nov. 2019.

[11] G. Kurniawan and Z. Rustam, “Enhancement of hepatitis virus outcome predictions with application of K-means clustering,” Proceedings of the 4th International Symposium on Current Progress in Mathematics and Sciences (ISCPMS2018), vol. 2168, no. 1, pp. 1–5, Nov. 2019.

[12] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, “An efficient k-means clustering algorithm: analysis and implementation,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pp. 881–892, Jul. 2002.