

XTERN Work Sample Data Science Report

Dataset

Complete Data

Location.Name	HousingDist	CoWork1Dist	CoWork2Dist	CoWork3Dist	CoWork4Dist	CoWork5Dist	type
Indy 500	4	9.4	13.8	19.8	7.7	7.1	sport
Talbot Street Art Fair	2.6	4.9	17.7	15.6	1.8	2.5	art
Indianapolis Indians Baseball Game	1.7	7.3	18	22.3	1.3	1	sport
5 Seconds of Summer Concert	1.8	7.4	18.1	22.4	1.3	1.3	music
Indianapolis CyberSecurity Conference	3.2	3.8	15.7	14.5	3.3	3.2	conference
Gen Con	2.2	7.3	18.5	22.8	1.1	0.8	gaming
Mama Carolla's	7.4	0.2	13.9	12.1	6.7	7.3	restaurant
The Jazz Kitchen	7.2	0.4	13.6	12.7	6.4	7	restaurant
Twenty Tap	7.2	0.4	13.6	12.4	6.4	7.1	restaurant
The Friendly Tavern	17.2	14.2	0.2	19.2	17.6	17.7	restaurant
Convivio Italian Artisan Cuisine - Zionsville	17.3	14.3	0.0714	19.4	17.8	17.9	restaurant
Delicia	6.9	0.4	13.9	12.4	6.1	6.8	restaurant
Diavola	7.6	0.3	14	12	6.8	7.4	restaurant
Fat Dan's Deli	7.2	0.4	13.6	12.4	6.4	7.1	restaurant
Cobblestone	17.2	14.2	0.1	19.3	17.7	17.8	restaurant
Chatham Tap Mass Ave	2.6	5.2	18.2	20.7	1.2	2.1	restaurant
IUPUI	0	7.4	18.1	22.4	2.7	2.4	housing

Sample 10 – Week Activities Plan Summer of 2022

Week 1 (May 29th – June 5th)

Weekly Dinner:

Mama Carolla's on Wednesday June 1st

Biweekly Event:

Indy 500

Details:

Date: Sunday May 29th

Location: 4790 W. 16th St., Indianapolis, IN 46222

Description: "The 1056th Running of the Indianapolis 500 presented by PennGrade Motor Oil will take place Sunday, May 29, 2021. Thirty-three drivers. 200 laps. 500 miles. One bottle of cold milk. The only thing missing from this magical Indianapolis 500 mix is you."

Category: "Sports"

Week 2 (June 5th – June 12th)

Weekly Dinner:

The Jazz Kitchen on Wednesday June 8th

Week 3 (June 12th – June 19th)

Weekly Dinner:

Twenty Tap on Wednesday June 15th

Biweekly Event:

Talbot Street Art Fair

Details:

Date: Saturday June 18th

Location: 16th and Talbott St. Indianapolis, IN 46202

Description: "With over 200 artists from across the nation, this juried art fair continues to be ranked as one of the finest fairs in the country. Talbot Street Art Fair is located between 16th & 20th / Delaware & Pennsylvania - Indianapolis in the historic Herron Morton Neighborhood."

Category: "Arts"

Week 4 (June 19th – June 26th)

Weekly Dinner:

The Friendly Tavern on Wednesday June 22nd

Week 5 (June 26th – July 3rd)

Weekly Dinner:

Convivio Italian Artisan Cuisine on Wednesday June 29th

Biweekly Event:

Indians Baseball Game

Details:

Date: Sunday June 26th

Location: 501 W Maryland St, Indianapolis, IN 46225

Description: "Indianapolis Indians play against Memphis Redbirds at Victory Field."

Category: "Sports"

Week 6 (July 3rd – July 10th)

Weekly Dinner:

Delicia on Wednesday July 6th**Week 7 (July 10th – July 17th)**

Weekly Dinner:

Diavola on Wednesday July 13th

Biweekly Event:

5 Seconds of Summer Concert

Details:

Date: Friday July 15th

Location: 801 W. Washington St. Indianapolis, IN 46204

Description: "5 Seconds of Summer (5SOS) are returning to the road this summer with the announcement of their No Shame Tour."

Category: "Music"

Week 8 (July 17th – July 24th)

Weekly Dinner:

Fat Dan's Deli on Tuesday July 19th

Conference:

Indianapolis Cyber Security Conference

Date: Wednesday July 20th

Location: 50 W Fall Creek Pkwy N Dr, Indianapolis, IN 46208

Description: "FutureCon Events brings high-level Cyber Security Training discovering cutting-edge security approaches, managing risk in the ever-changing threat of the cybersecurity workforce. 'Cybersecurity is no longer just an IT problem'"

Category: "Conference"

Week 9 (July 24th – July 31st)

Weekly Dinner:

Cobblestone on Wednesday July 27th

Biweekly Event:

Gen Con

Details:

Date: Friday July 29th

Location: 100 S. Capitol Ave., Indianapolis, IN 46225

Description: "Gen Con Indy is the original, longest-running, best-attended, gaming convention in the world. This year's event offers concurrent in person and online experiences: Gen Con Indy, Gen Con Online, and Pop-Up Gen Con."

Category: "Gaming"

Week 10 (July 31st – August 7th)

Weekly Dinner:

Chatham Tap Mass Ave on Wednesday August 3rd

Explanation for the 10 – Week Sample Plan

Restaurants:

The restaurants for the 10-week sample plan were strategically chosen. First, using Google APIs the top 20 restaurants near each coworking location were collected. Then these restaurants were added to a along with other information about them obtained from the API such as rating and price level. Next, the table was read into R and the restaurants with price levels that were too high or too low were removed. This was done because it is unrealistic that weekly group dinners would be at 5-star, expensive restaurants, but it also is expected that weekly group dinners will be at nicer dinners than subway and McDonald's.

After removing these rows, the number of restaurants was reduced, but not enough. Therefore, the data was then sorted by rating, and the top 20 restaurants were selected. After this, the top 10 restaurants were chosen based on research using google and yelp, depending on criteria such as when the restaurant is open (some were breakfast restaurants) and the type of food served there. Finally, the top 10 restaurants were selected.

Events:

The events were much more difficult to choose, as there are not that many events posted yet for May-August of 2022. A reconsideration of the events in the spring (March or April) will likely yield better results, and it is highly recommended that this is done

All the biweekly events were selected because of the opportunity for social interaction. Group members/coworkers will be able to talk and get to know each other at these events such as the Indy 500 or the 5 Seconds of Summer concert. This kind of relationship building would not be possible at if events such as movies or plays were selected where it is expected that participants simply go and sit to watch the show quietly.

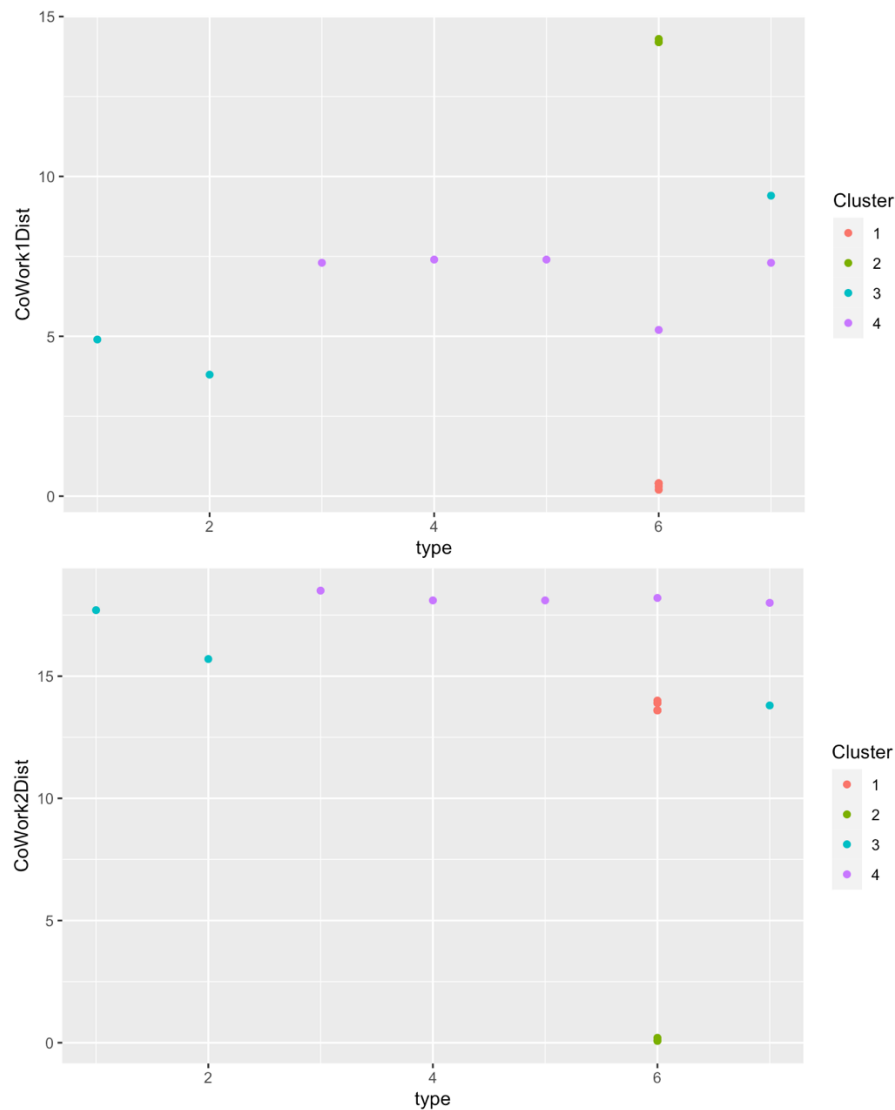
The events were selected to represent top things to do in Indy. The reason for this is to expose XTERNs, especially from out of state to how great the city of Indianapolis is. The events were also chosen to appeal to a variety of interest. Two “sporting” events were selected: The Indy 500 and an Indianapolis Indians baseball game because these are good representations of sports in Indy during the summer. The art fair was selected to appeal to individuals with artistic interest and is an opportunity that will allow all XTERNs to learn more about the art and culture in Indiana while enjoying food and the company of their group. The 5 Seconds of Summer concert was selected because this is a band that many young adults listen to, and generally concerts at the White River State Amphitheatre are much more affordable. Lastly, Gen Con was selected to appeal to XTERNs that like to game, as this is one of the biggest gaming events that happens in Indianapolis every year.

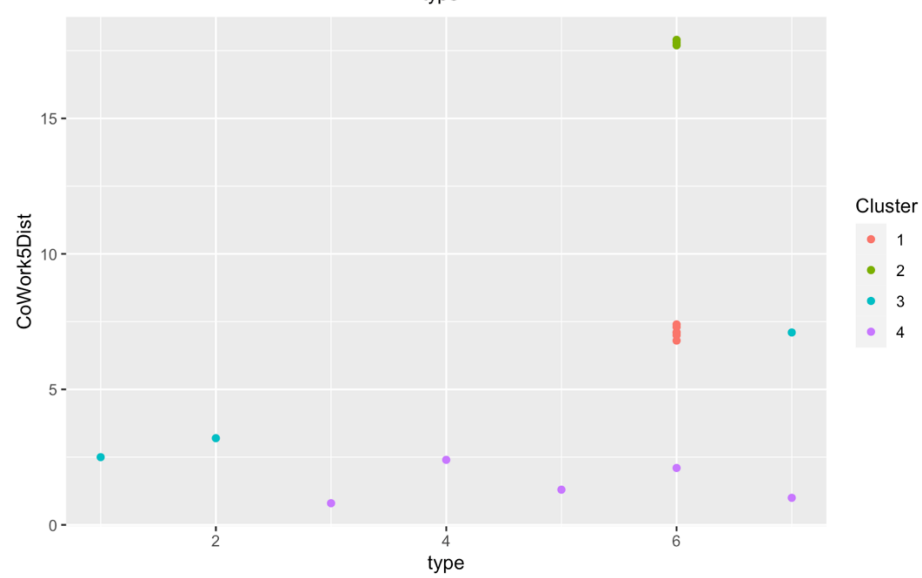
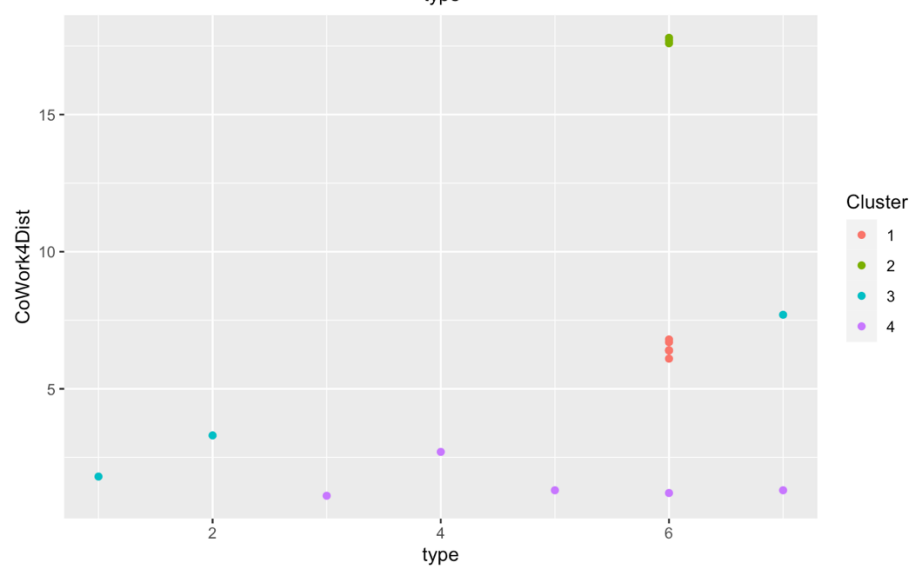
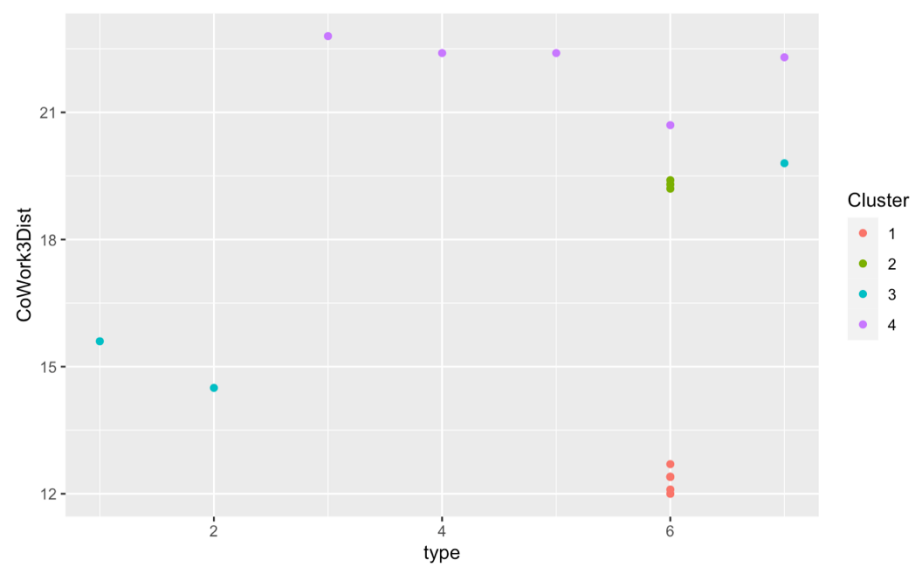
Finally, there were extremely limited options for conferences that have already been posted for next summer. For this reason, it is suggested that the conference choice is reevaluated in the spring months. However, under the circumstances, the Indianapolis Cyber Security conference was selected as the big conference/talk activity for the summer. It was selected because the individuals participating in the XTERN program are all interested in the STEM field and Cyber Security is consistently becoming more important to this field. So, this would be a conference that all individuals would learn and benefit from.

Cluster Analysis

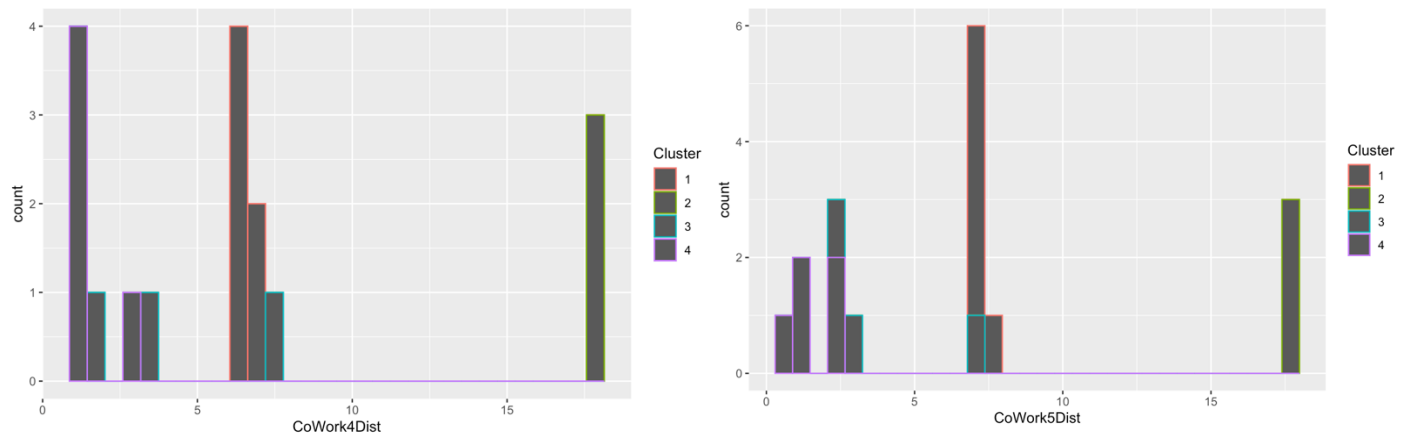
After creating the data table, the next step was to cluster the data to determine the most convenient co-working location. Kmeans clustering algorithm was used to perform cluster analysis on this dataset. This algorithm resulted in 4 clusters which are shown below in the graphs by the four colors.

Below are some data visualizations showing these clusters. The graphs below show the type of location on the x-axis (type 1 = art, type 2 = Conference, type 3 = Gaming, type 4 = Housing, type 5 = Music, type 6 = Restaurant, type 7 = sport) and the distance from the various coworking locations on the y-axis. The graphs also show the clusters that each data point belongs to depending on their color.





There are several key takeaways from the graphical visualizations above. First, ignoring the clusters/individual colors we can simply evaluate the convenience of the coworking location based on how far away the individual points are from the x-axis. Looking at the first graph we see that most points are a moderate distance from coworking location 1 (5-10 miles). This implies that this coworking location could be a good fit. Next, in graphs 2 and 3 that many of the event and restaurant locations are much further away from coworking location 2 and 3 respectively indicating that these locations would be less convenient. Finally, looking at graphs 4 and 5 we see that most of the event locations are close to (within 7 miles) the coworking locations. This implies that coworking location 4 or 5 is the most convenient. Let's look at another visualization to determine if location 4 or 5 is the most convenient.



The above visualizations show the number of that occur at each distance from the coworking location. From this we can see that both location 4 and 5 have most of their events occurring within 7 miles of the coworking location. However, it appears that location 4 is slightly preferable, because there are 5 events within 2 miles of the coworking location and looking closely more events are slightly closer to location 4 than to location 5. So, in conclusion from the above graphs coworking location 4, Industrious Mass Ave, is the most convenient coworking location given this 10-week sample activity plan and location 5, Launch Indy, is a close second.

These insights could have been obtained without the cluster analysis. However, the clustering of these events gives further insights. For example, cluster 1 contains only restaurants and these restaurants are close or moderately close to all coworking locations except for location 2. Cluster 2 also contains only restaurants, and these tend to be far away from every coworking location except for location 2. Cluster 3 contains a mix of events, but no restaurants and these events are a close to moderate distance from every coworking location except for coworking location 2. Finally, Cluster 4 contains a variety of events and restaurants and these all appear to be very close to coworking location 4 and 5, but far from location 2 and 3.

Overall, this Cluster Analysis helps us to gain a better understanding of the data, but in the end the distance of the events from the various coworking locations is more meaningful than the kmeans clustering, given this specific situation.

Conclusions

This dataset was created with location and distance as the most important factors. The restaurants were chosen based on their proximity to the various coworking spaces, and the distance of the activities from each colocation were collected. Given more time it would be important to collect other data and perhaps other criteria such as reviews and event venue size as variables in our dataset. Also, given more time, or doing this project in the spring would result in more event/activities to choose from to ensure the very best activities are being chosen for the XTern program. It is important to note that changing the 10-Week sample plan could completely change which colocation is most convenient. However, given the sample plan in this report, it was found that coworking location 4, Industrious Mass Ave, would be the most convenient location.

Resources

<https://www.visitindy.com/indianapolis-top-events>

<https://www.visitindy.com/indianapolis-gaming>

<https://www.visitindy.com/indianapolis-guides>

<https://futureconevents.com/events/indianapolis-in-2022/>

<https://img.mlbstatic.com/mlb-images/image/upload/milb/eqe0bqa8nf0yrjvxkjp.pdf>

<https://www.vividseats.com/indianapolis-indians-tickets-victory-field-6-26-2022--sports-minor-league-baseball/production/3746868>