**UNINORTE**

**Doctorado en Ingenieria Industrial**

**Minería de Datos**

**Quiz #4**

**Instructions:**

* Present your Quiz #4 using Word, PDF, Shiny, R Markdown or other related.
* Send your files to [silvasotillo@usf.edu](mailto:silvasotillo@usf.edu)
* Deadline for submission is April 25th at 11:59pm

**Case 1**

Steps to follow:

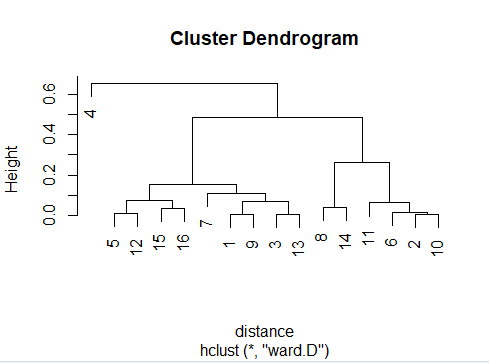
1. Data presents reviews from restaurants (see Data\_yelp.xlsx).
2. There are 500 positive sentences and 500 negative sentences (sentiment labelled).
3. Apply Text Analytics to analyze this data set.
4. Define your output variable, the objective of the study, assumptions and all what you consider necessary.
5. Your report should include the Sections that you best consider. It also has to include as much detail as you can for all steps followed. Include tables and/or graphs to visualize each partial result during your data analysis process.
6. You may include visualizations tools to better present your findings.
7. You should submit all files needed to allow the instructor to reproduce your work

**Case 2**

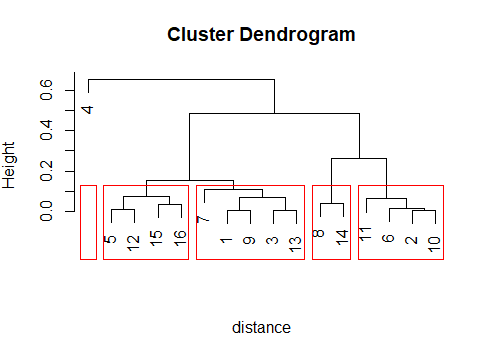
1. All students (16) will share their addresses (or one of interest). Use any electronic form for sharing.

| Tag | Nombre | Lat | Lon |
| --- | --- | --- | --- |
| 1 | Cesar Pajaro | 11.007 | -74.834 |
| 2 | Danny Daniel Ortega | 10.982 | -74.809 |
| 3 | Efrain Boom carcamo | 10.992 | -74.845 |
| 4 | Jose Navarro De la Cruz | 10.788 | -74.755 |
| 5 | Joseph Soto | 11.019 | -74.815 |
| 6 | Salvador Villamizar | 10.983 | -74.804 |
| 7 | Maria Isabel Arrieta Escobar | 11.026 | -74.865 |
| 8 | Nazhir Amaya | 10.902 | -74.787 |
| 9 | Jhoan Castro | 11.006 | -74.835 |
| 10 | Loraynne Amaya | 10.977 | -74.809 |
| 11 | Valentina Mejia | 10.963 | -74.795 |
| 12 | Ethel Garcia | 11.017 | -74.818 |
| 13 | Kevin Palomino | 10.993 | -74.847 |
| 14 | Carlos Ferreira | 10.915 | -74.801 |
| 15 | Fernando Gonzalez | 10.996 | -74.822 |
| 16 | Ana Luisa Cuello | 10.997 | -74.802 |

1. Present a dendrogram (use Manhattan distances).



1. Identify the optimal number of clusters



1. Use k-means algorithm to divide data into the number of clusters found in the previous step.

| Tag | Nombre | Lat | Lon | Cluster | Lat\_Center | Long\_Center | DistCenter |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | Cesar Pajaro | 11.007 | -74.834 | 1 | 11.004 | -74.831 | 19.38 |
| 2 | Danny Daniel Ortega | 10.982 | -74.809 | 4 | 10.980 | -74.804 | 8.42 |
| 3 | Efrain Boom carcamo | 10.992 | -74.845 | 1 | 11.004 | -74.831 | 76.90 |
| 4 | Jose Navarro De la Cruz | 10.788 | -74.755 | 2 | 10.788 | -74.755 | 0.00 |
| 5 | Joseph Soto | 11.019 | -74.815 | 1 | 11.004 | -74.831 | 94.34 |
| 6 | Salvador Villamizar | 10.983 | -74.804 | 4 | 10.980 | -74.804 | 15.58 |
| 7 | Maria Isabel Arrieta Escobar | 11.026 | -74.865 | 3 | 11.026 | -74.865 | 0.00 |
| 8 | Nazhir Amaya | 10.902 | -74.787 | 5 | 10.909 | -74.794 | 41.57 |
| 9 | Jhoan Castro | 11.006 | -74.835 | 1 | 11.004 | -74.831 | 8.38 |
| 10 | Loraynne Amaya | 10.977 | -74.809 | 4 | 10.980 | -74.804 | 19.83 |
| 11 | Valentina Mejia | 10.963 | -74.795 | 4 | 10.980 | -74.804 | 108.60 |
| 12 | Ethel Garcia | 11.017 | -74.818 | 1 | 11.004 | -74.831 | 78.49 |
| 13 | Kevin Palomino | 10.993 | -74.847 | 1 | 11.004 | -74.831 | 69.06 |
| 14 | Carlos Ferreira | 10.915 | -74.801 | 5 | 10.909 | -74.794 | 41.57 |
| 15 | Fernando Gonzalez | 10.996 | -74.822 | 1 | 11.004 | -74.831 | 54.64 |
| 16 | Ana Luisa Cuello | 10.997 | -74.802 | 4 | 10.980 | -74.804 | 104.43 |

Map

Description automatically generated

Your final report should have no less than 5 pages and no more than 10 pages

Good luck !