# **CSE 4062 - Semester 2021 Intro to Data Science and Analytics Delivery #5 - Descriptive Analytics**



## **Group 2 Members:**

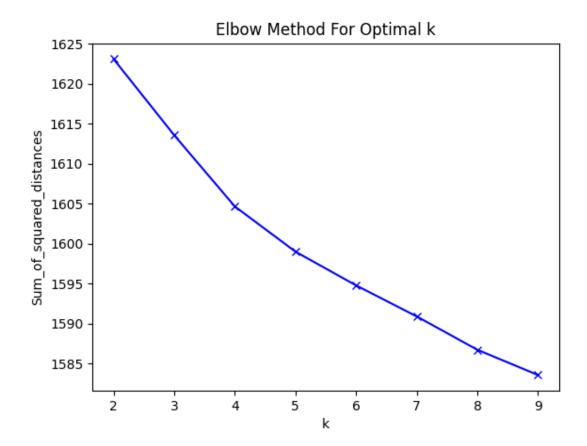
Aleyna BOZACI 150319630 - aleynabozacii@gmail.com Bekir ÖZKAN 150319557 - bekirozkan9698@gmail.com Merve YAYIN 150116051 - yayinm8@gmail.com Metehan ERTAN 150117051 - metehan.ertan@hotmail.com Mustafa USCA 150416054 - uscamustafa64@gmail.com

## 1 - Statistics

To find optimal K for k-means clustering we used the Elbow method. After finding the optimal K we created clusters and displayed most used words with WordCloud. As our datasets consist of the same topic most of the clusters got the same most used words. Most used ones change after the third ones.

## 1.1 - Yelp dataset

## 1.1.1 - Clustering Elbow method



## 1.1.2 - Clusters

id	cluster
1495	0
651	0
1459	0
1386	0
289	0
1318	6
1318 1319	6 6
1319	6

[1800 rows x 2 columns]

## 1.1.3 - Cluster 0



## 1.1.4 - Cluster 1



## 1.1.5 - Cluster 2



## 1.1.6 - Cluster 3

```
topping oven pizzamuch w come finargherita almays back oven best back oven best best pizza oven best pizza best pizza place price crustopy perfect over best pizza place price crustopy perfect over best pizza place price pr
```

## 1.1.7 - Cluster 4



## 1.1.8 - Cluster 5

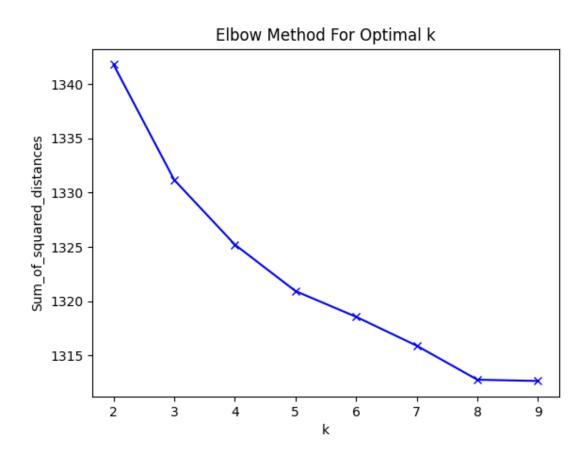


## 1.1.9 - Cluster 6



## 1.2 - Turk dataset

## 1.2.1 - Clustering Elbow method



## **1.2.2** - Clusters

id	cluster
1599	0
1087	0
1090	0
1091	0
1096	0
	•••

```
1358 6
1355 6
994 6
1352 6
1125 6
```

[1600 rows x 2 columns]

## 1.2.3 - Cluster 0



#### 1.2.4 - Cluster 1

```
clean year well will friendly re work place of say nice place of s
```

#### 1.2.5 - Cluster 2

```
price place in well of location day work small lathough of guest disappointed restaurant much bathroom of service chicago better towel one time took window will comfortable even took window window took window think took window think took one time took window will comfortable even took one took window took window took window took window took shower shower saydon t great overall
```

#### 1.2.6 - Cluster 3

```
thing call chicago went asked still experience bad smything charge use took seemed tooked front desk left said ready two someone really experience bad smything charge use took took two seemed tooked front desk left said ready two someone really entired in light reservation rude way arrived finally stay with two bed never better wasn t bathroom take wait stayed another check nice wasn t bathroom wasn to the companion of the comp
```

#### 1.2.7 - Cluster 4

```
hour placerestaurant room service minute of the commend of the com
```

## 1.2.8 - Cluster 5



## 1.2.9 - Cluster 6



## 2 - Description of results

As mentioned before our datasets mostly focus on one topic. Because of this each cluster has some same words. For each cluster most used words start with the same 2 words. After that each cluster focusses on another topic. Clustering method is not really efficient for our project and datasets.