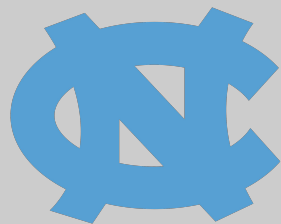


# Datafest 2018

Group jobs by common characteristics shared by these postings

---



Weibin's Warriors

April 8th, 2018

# Methodology

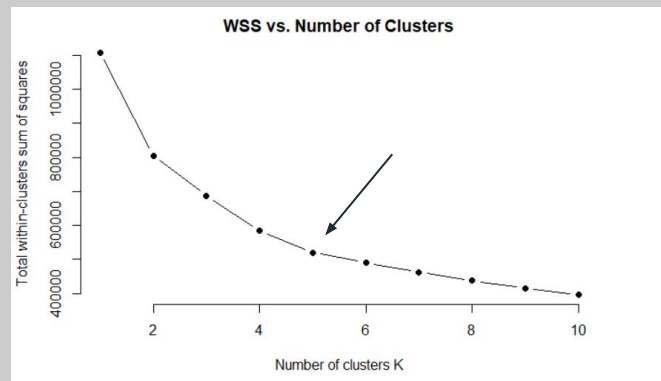
---

## Variables:

- Job ID
- Salary
  - Factored in cost of living
- U.S. states
- Job Category
- Season posted
- Word count
- Clicks (per day)

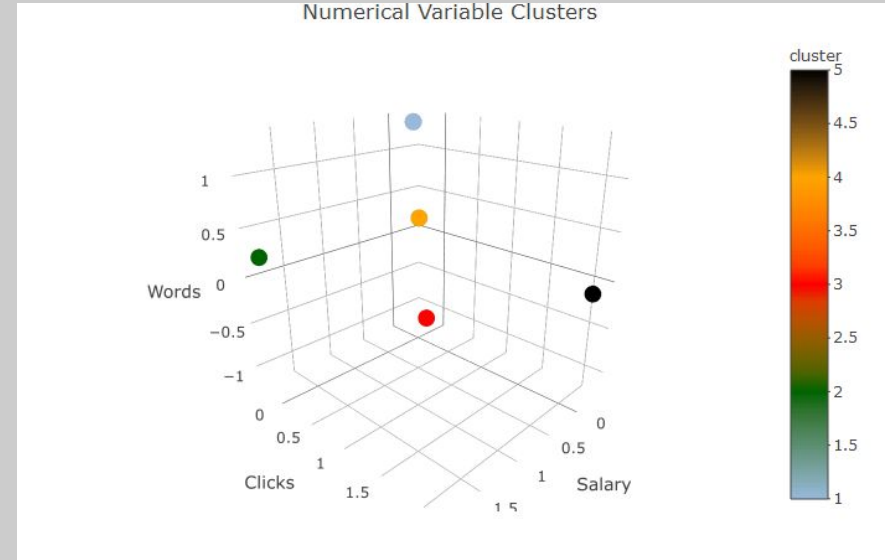
## Steps:

- NA reassignment
- Outlier reassignment
- Aggregation by JobID
- Variable selection
- Variable transformation
- Variable standardization
- **K-means clustering:**
  - Used numeric variables
  - Opt. amount of clusters with “elbow curve”



# Results

Cluster	Characteristic	Most Common	Least Common
1	High Word Count	Service	Tech, medicine
2	High Salary	Tech, engineering, medicine	Food, personal, retail, warehouse
3	Low Word Count	Veterinary, sanitation, personal, science	Finance, marketing, military, socialscience
4	Average	Admin, agriculture, service	Childcare, engineering, technology
5	High Clicks Per Day	Service, tech software	agriculture, military, Chemical engineering,



Interesting Fact: Location showed up significant only for cluster 2 - NC is the best apparently