Cluster Analysis: Inpatient DRG Costs

Group 5: Erin Cao, Henry Gao, Qi He, Yu Hsiang Hung, Robert Malongo, and Jiajie Yuan
Date : December 4, 2019

DRGs

DRG: A patient classification system that standardizes prospective payment to hospitals - categorizes inpatient hospitalization costs

DRG Parameters

- Principal diagnosis
- Secondary diagnosis(es)
- Surgical procedures performed
- Concurrent illnesses and complications
- Patient's age and sex

K-Means Clustering: Introduction

- Unsupervised machine learning algorithm
- The algorithm will categorize the items into K groups of similarity
- To calculate that similarity, we will use the euclidean distance as measurement

The algorithm works as follows:

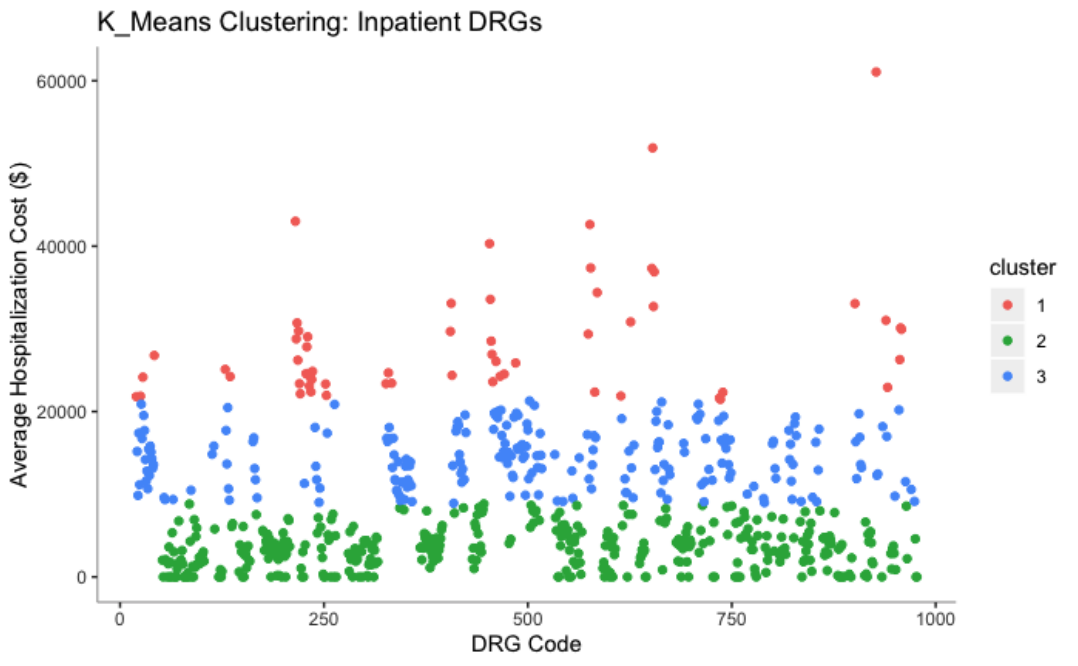
1. First we initialize k points, called means, randomly.
2. We categorize each item to its closest mean and we update the mean's coordinates, which are the averages of the items categorized in that mean
3. We repeat the process for a given number of iterations and at the end, we have our clusters.

K-Means Cluster Analysis: Method

Important Data Cleaning Steps

- Choosing DRGs between 20 and 977 only
- Dropping revenue charges below $100
- Filtering out all PCCR cost category
- Combining PCCR3700 and PCCR4000 into one new cost category
- Replacing all empty cells with value 0

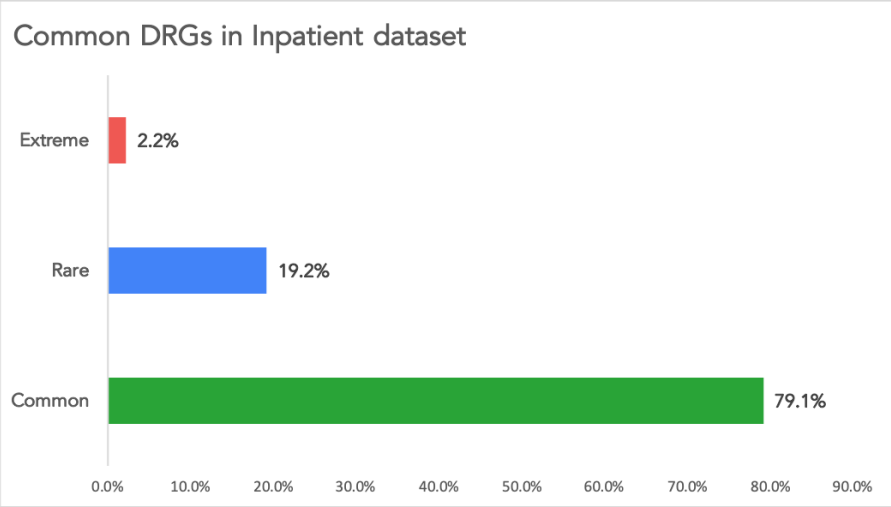
K-Means Cluster Analysis: Results, K=3

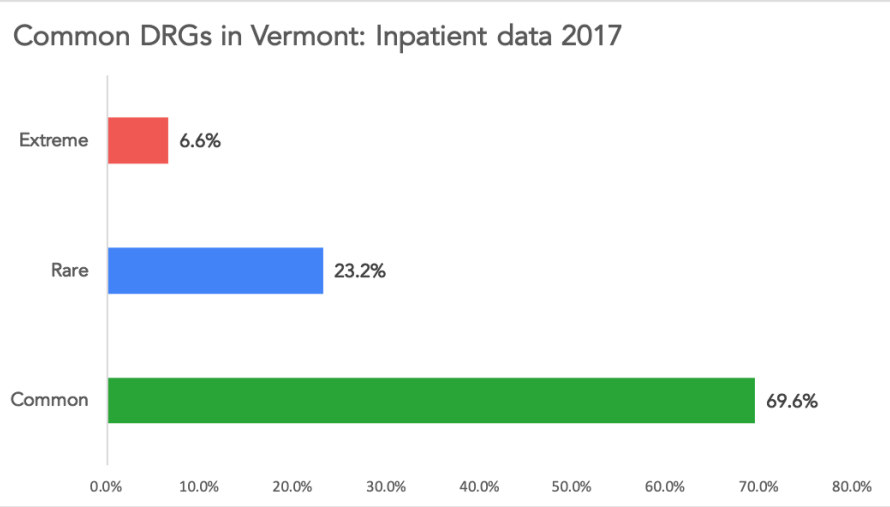


- **Red**: Extreme conditions
- **Blue**: Rare conditions
- **Green:** Common conditions

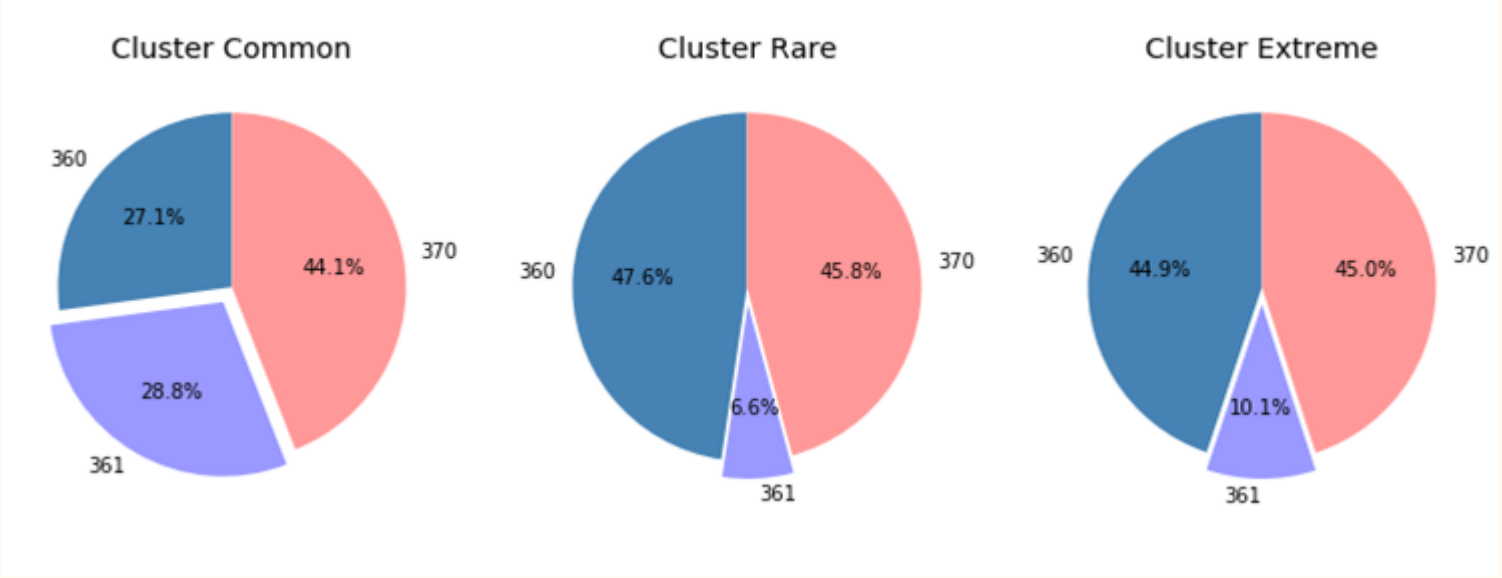
Frequency of DRG Conditions

- High number of patients in Common Cluster
- Common DRGs in Vermont





Type of Service



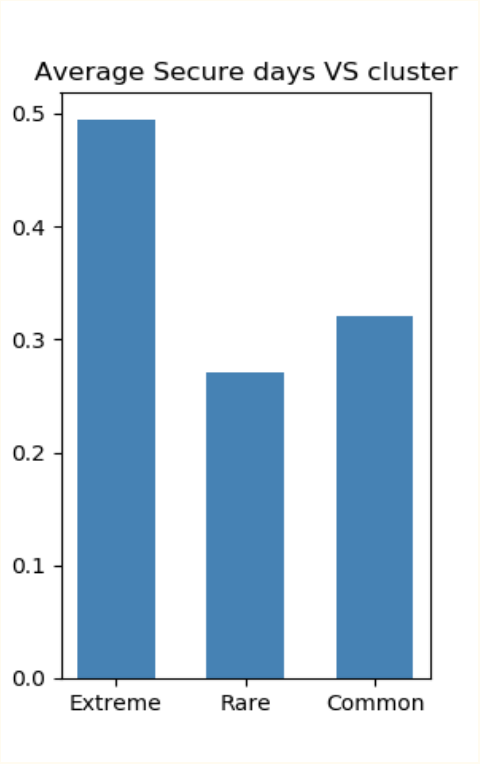
360 Operating Room Services

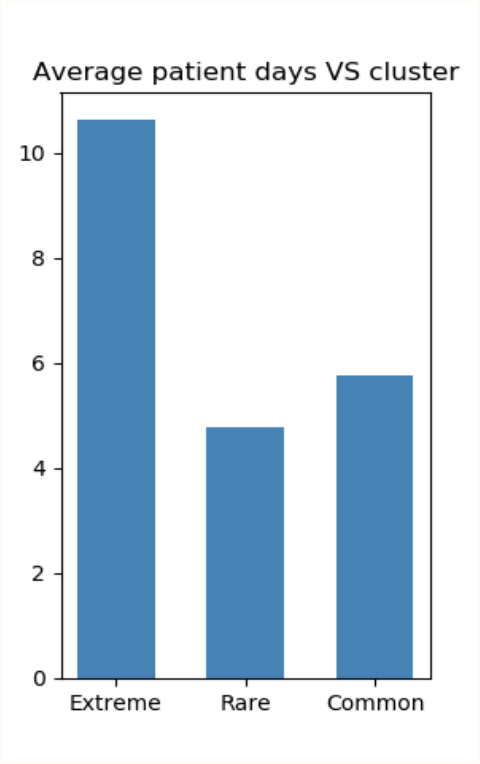
361 Operating Room Services: Minor surgery

370 Anesthesia

The portion of service provided (Revcode) among three clusters

Length of Stay in Hospital





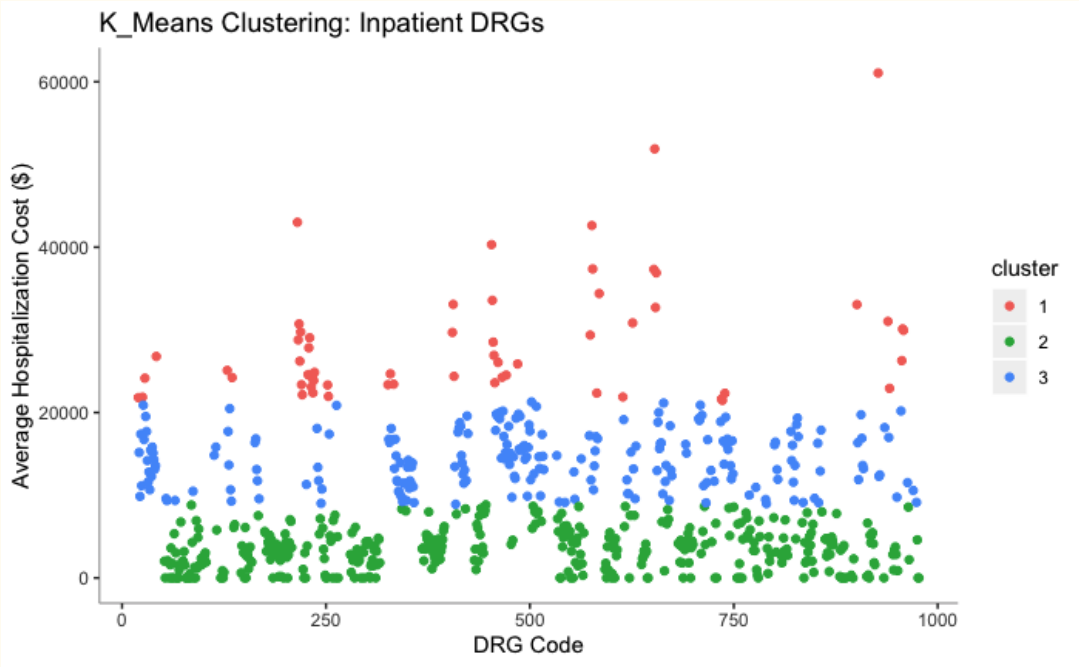
Secure days: Number of days in special care unit

Patient days: Total number of days in the hospital

T-test results

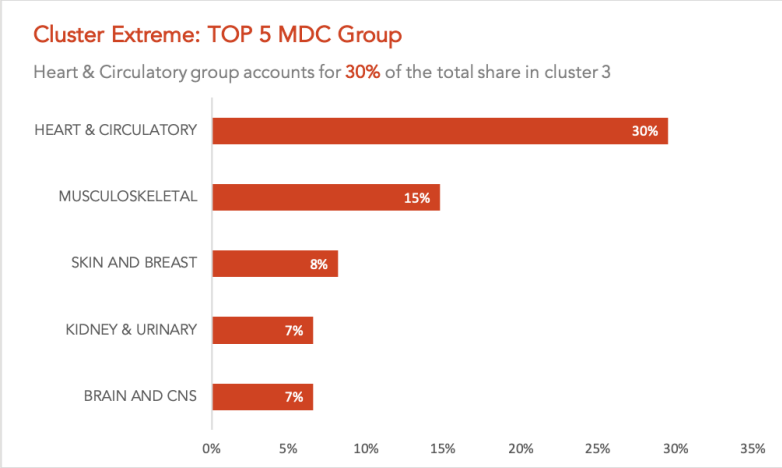
Difference between "Extreme" and "Rare"	**P-value**
Average secure days	0.004
Average patient days	~0

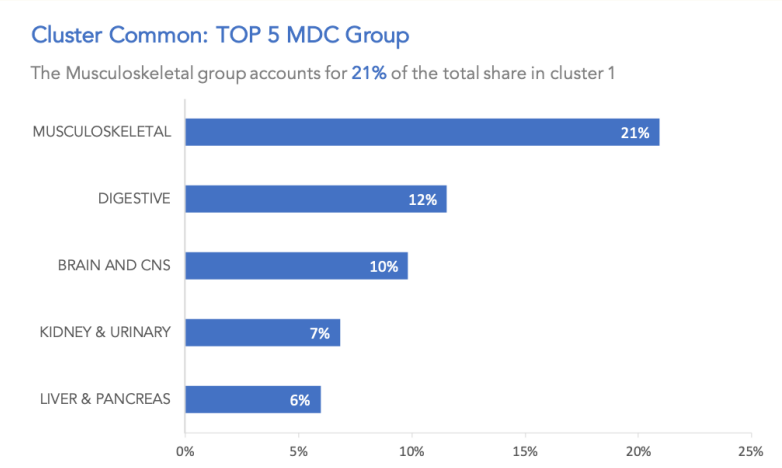
K-Means Cluster Analysis: Results, K=3



- **Red**: Extreme conditions
- **Blue**: Rare conditions
- **Green:** Common conditions

Top 5 MDC Group in Extreme vs. Common Cluster





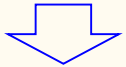
- Heart assist system implant
- skin grafting
- Cardiac valve catheterization
- Other cardiothoracic procedures

- Infections
- Respiratory
- Wound debridement

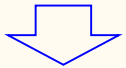
Conclusion

Cluster

Frequency of DRGs Conditions



Type of service



Length of patient stay in hospital

Top 5 MDC group in different clusters

Findings:

1. More patients are in hospital due to common conditions, less are there because of extreme conditions
2. Extreme conditions cost patients more on average (typically over $20,000)
3. The cost variance within extreme conditions is much larger (up to $60,000)
4. Patients with common conditions have more minor surgery
5. Patients with extreme conditions tend to stay in hospital for a longer time

Recommendation:

- Group patients by their conditions, set up different insurance plans
- When conducting medical background check, focus on the top areas in extreme conditions such as heart disease
- Referral Hospitals should prepare more rooms for patients