# OPM 562
# Case study:

## Unsupervised learning for failure mode detection

Son Phan, 1713240

04.05.2020

# Contents

# 1. Data cleaning & application of ML methods

Data preparation

Before further analysis, 24 rows were dropped from the dataset:

- 19 rows with missing data
- 5 rows with negative or zero values

To get deeper insights into the features, the dataset was divided into 2 parts:

- Frame data (top tube, seat tube, chainstay, front center)
- Battery data (weight, voltage, capacity)

ML methods

Combination of PCA and K-Means clustering to improve segmentation results:

- Dimension reduction prior to data segmentation
- PCA helps reduce noise and multicollinearity, making features both statistically significant as well as non-correlated with each other
- K-Means clustering method uses Euclidean distance as measurement of within cluster distance but does not work well in high dimensions
- We reduce the number of dimensions first so that K-means method produces more meaningful results

# 2.1 Frame Analysis

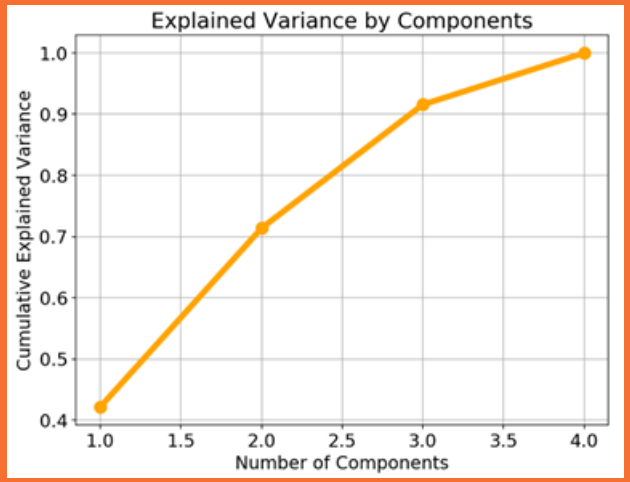
## 2.1.1 PCA & K-Means

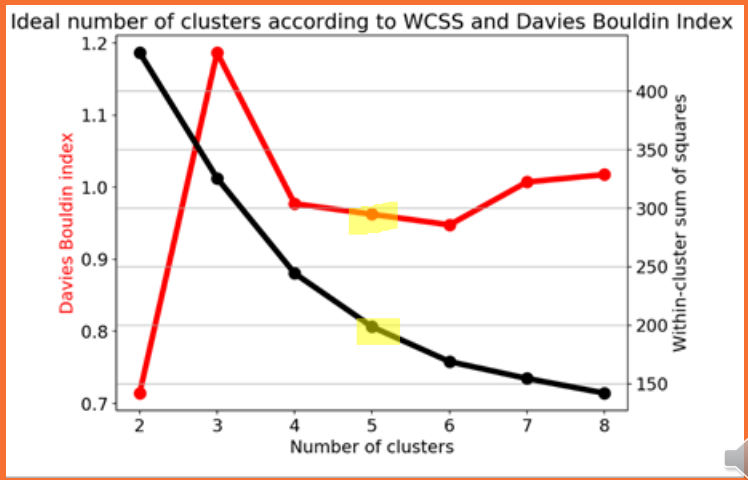
### Principal Component Analysis

- Original number of components: 4
- By reducing to 3 components, we are still able to explain 90% of the variance
- Thus, adjusted number of components: 3

### K-Means Clustering

- By looking at the Elbow graph (black), it can be noted that the point of inflection occurs approximately around the 5th cluster.
- By choosing 5 clusters, we get a Davies Bouldin Index of 0.9476
- Thus, number of clusters: 5

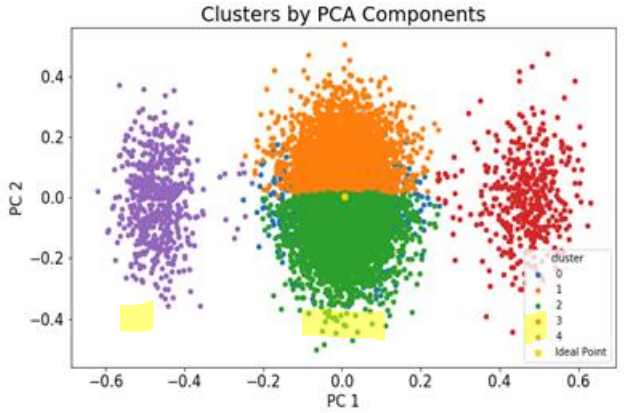


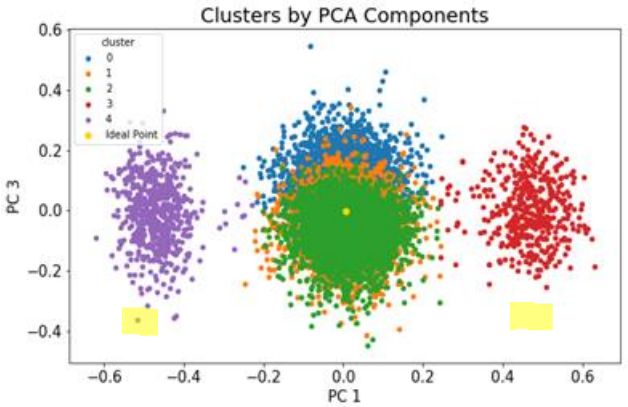
# 2.1 Frame Analysis

## 2.1.2 Clustering

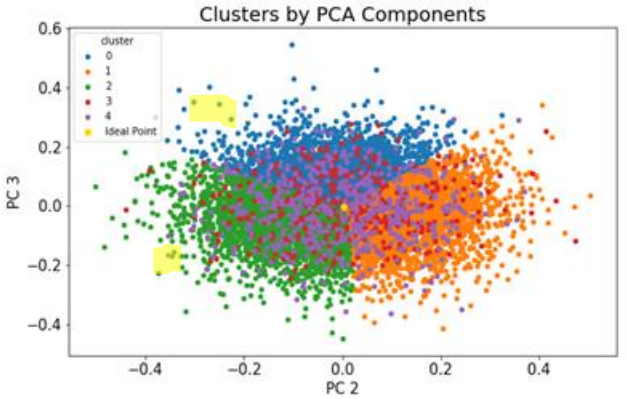
Bike frame with ideal features (nominal values) belongs to cluster 0




- Clusters 3 and 4 clearly separated with respect to PC 1
- PC 1 is mainly explained by top tube and front center




- Clusters 1 and 2 determined by PC 2
- PC 2 is mainly explained by chainstay



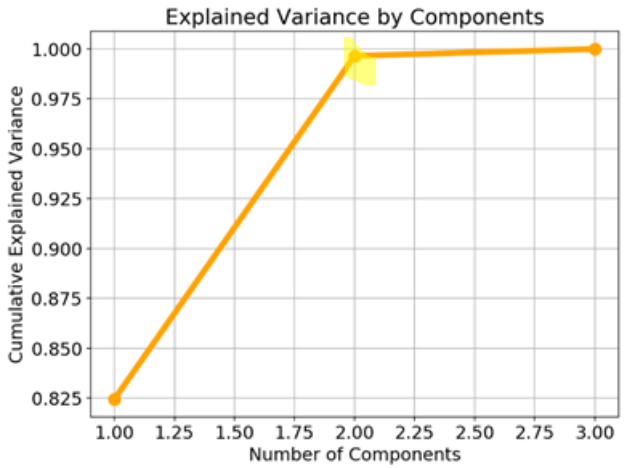

- Clusters 0 and 2 determined by PC 3
- PC 3 is mainly explained by seat tube

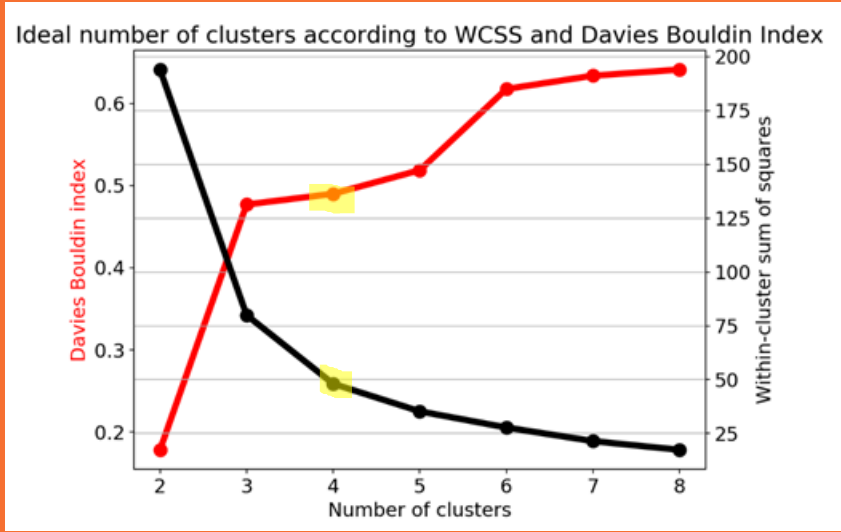
# 2.1 Frame Analysis

## 2.1.3 Deviation from nominal values

- Low deviations from nominal value
- Medium-sized deviations
- Large deviations

| | Nominal value (in mm) | Cluster 0 28.24% | Cluster 1 31.78% | Cluster 2 30.2% | Cluster 3 4.3% | Cluster 4 5.48% |
|---|---|---|---|---|---|---|
| Top tube | 602 | 601.998879 -0,001121 | 602.021299 +0,021299 | 601.994177 -0,005823 | 600.234905 -1,765095 | 604.624928 +2,624928 |
| Seat tube | 584 | 584.962837 +0,962837 | 583.602228 -0,397772 | 583.555549 -0,444451 | 583.922130 -0,077870 | 584.051143 +0,051143 |
| Chainstay | 487 | 486.889788 -0,110212 | 487.935831 +0,935831 | 486.045182 -0,954818 | 486.940714 -0,059286 | 486.970386 -0,029614 |
| Front center | 633 | 632.988222 -0,011778 | 632.997101 -0,002899 | 632.991718 -0,008282 | 636.775588 +3,775588 | 630.257684 -2,742316 |




# 2.2 Battery Analysis

## 2.2.1 PCA & K-Means

### Principal Component Analysis

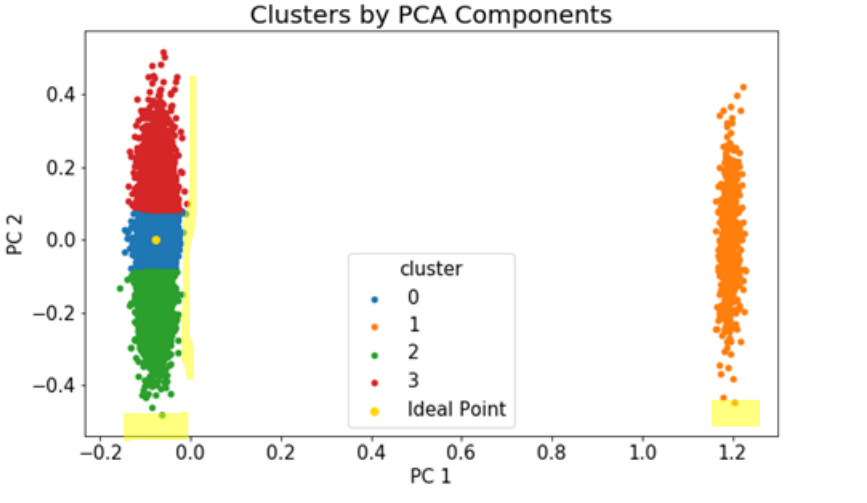
- Original number of components: 3
- By reducing to 2 components, we are still able to explain almost 100% of the variance
- Thus, adjusted number of components: 2

### K-Means Clustering

- By looking at the Elbow graph (black), it can be noted that the point of inflection occurs approximately around the 4th cluster.
- By choosing 4 clusters, we get a Davies Bouldin Index of 0.5188

# 2.2 Battery Analysis

## 2.2.2 Clustering




- Battery with ideal features (nominal values) belongs to cluster 0
- Two clearly separated clusters with respect to PC 1
- Clusters on the left around the ideal point, cluster on the right much higher than ideal point.
- PC1 is mainly explained by voltage and capacity
- Clusters 0, 2 and 3 differ with respect to PC 2
- PC 2 is mainly explained by battery weight

# 2.1 Battery Analysis

## 2.2.3 Deviation from nominal values

- Low deviations from nominal value
- Medium-sized deviations
- Large deviations

| | Nominal value | Cluster 0<br>44.2% | Cluster 1<br>6.04% | Cluster 2<br>24.16% | Cluster 3<br>25.59% |
|---|---|---|---|---|---|
| Battery weight | 2.8 kg | 2.800075 kg<br>+ 0.00075 kg | 2.799693 kg<br>- 0.000307 kg | 2.805238 kg<br>+ 0.005238 kg | 2.794915 kg<br>- 0.005085 kg |
| Full charge voltage | 400 Wh | 400.006 Wh<br>+ 0.006 Wh | 281.245 Wh<br>-118.755 Wh | 400.002 Wh<br>+ 0.002 Wh | 399.982 Wh<br>- 0.018 Wh |
| Battery capacity | 36 V | 36.000318 V<br>+ 0.000318 V | 34.198396 V<br>- 1.801604 V | 36.001027 V<br>+ 0.001027 V | 35.998603 V<br>- 0.001397 V |

# 3. Managerial implications

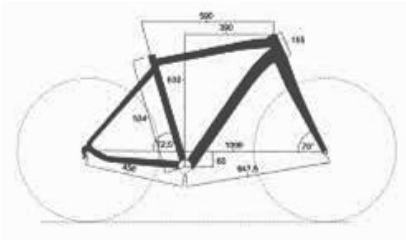
## Quality of frames

**Reasonable:**
**28%** of frames (cl. 0)

**Questionable:**
**62%** of frames (cl. 1&2)

**Poor:**
**10%** of frames (cl. 3&4)

| Acceptable deviations from the nominal values, except for rather long seat tube | Somewhat shorter seat tube & noticeably longer or shorter chainstay | Very short top tube and very long front center or vice versa |
|---|---|---|



Top tube and front center are inversely proportional, which in particular leads to 2 clusters of poor quality.

From the bike scheme, it seems like these deficiencies result from inaccurate connection of the tubes. It can result from not properly set up machines and/or too rare maintenance.

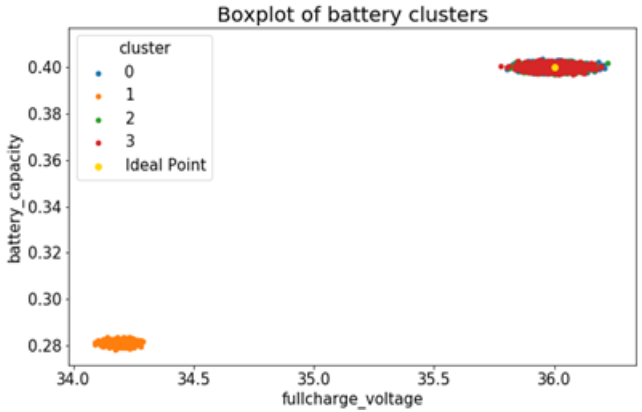
# 3. Managerial implications

## Quality of batteries

**44%** batteries (cl. 0) - **good** quality with acceptable deviations from nominal values.

**50%** batteries (cl. 2&3) - **reasonable** quality with rather low/high weight.

**6%** batteries (cl. 1) - **poor** quality with very low capacity and very low full charge voltage.




## Overall conclusions and suggestions

- **27%** of bikes - good or reasonable quality (over both parts)
- **69%** of bikes have issues with either frame or battery
- **4%** of questionable or poor quality bikes (over both parts)

| Parts produced by Sunflower Bikes | | Parts bought from a supplier | |
|---|---|---|---|
| Analyse machines data, investigate causes of defects | Adjust schedules of maintenance, improve accuracy of machines | Communicate the problem to the supplier | Implement a stricter sampling policy, reject shipments with high % of defects |

# Thank you