**Files for clusters**

For top performers and middle perfomers respectively, I put the files in Github: (Emily/Files\_for\_clusters)

1) Files:

**dayperiod\_violationcategory\_forcluster\_middle\_performers**

**dayperiod\_violationcategory\_forcluster\_top\_performers**

Feature1: day\_period = 2 possibilities (AM/PM)

Feature2: violation\_category = 6 possibilities (1 to 6)

Equals = 12 combined features (ex AM\_1)

2) Files:

**dayperiod\_violationcode\_forcluster\_middle\_performers**

**dayperiod\_violationcode\_forcluster\_top\_performers**

Feature1: day\_period = 2 possibilities (AM/PM)

Feature2: violation\_code = 99 possibilities (1 to 99)

Equals = 198 combined features (ex AM\_50)

3) Files:

**dayweek\_violationcategory\_forcluster\_middle\_performers**

**dayweek\_violationcategory\_forcluster\_top\_performers**

Feature1: day\_week = 7 possibilities (1 to 7, 1= Monday, 7 = Sunday)

Feature2: violation\_category = 6 possibilities (1 to 6)

Equals = 42 combined features (ex 3\_50)

4) Files:

**dayweek\_violationcode\_forcluster\_middle\_performers**

**dayweek\_violationcode\_forcluster\_top\_performers**

Feature1: day\_week = 7 possibilities (1 to 7, 1= Monday, 7 = Sunday)

Feature2: violation\_code = 99 possibilities (1 to 99)

Equals = 693 combined features (ex 3\_50)

5) Files:

**typeday\_violationcategory\_forcluster\_middle\_performers**

**typeday\_violationcategory\_forcluster\_top\_performers**

Feature1: type\_day = 2 possibilities (Week, Weekend)

Feature2: violation\_category = 6 possibilities (1 to 6)

Equals = 12 combined features (ex Week\_3)

6) Files:

**typeday\_violationcode\_forcluster\_middle\_performers**

**typeday\_violationcode\_forcluster\_top\_performers**

Feature1: type\_day = 2 possibilities (Week, Weekend)

Feature2: violation\_code = 99 possibilities (1 to 99)

Equals = 198 combined features (ex Week\_46)