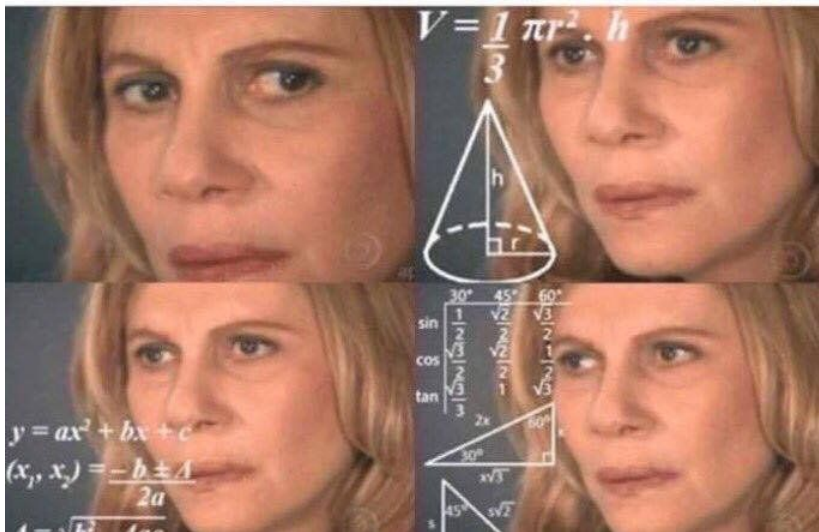




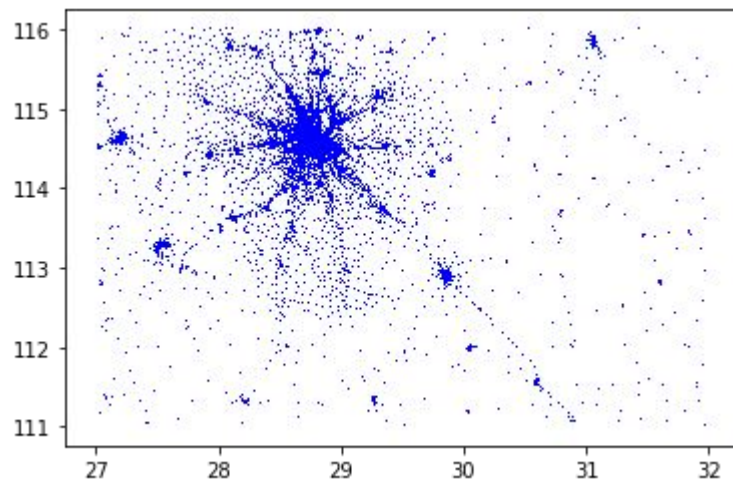
# CUP\_IT\_2020 DATA SCIENCE

Команда Пирог



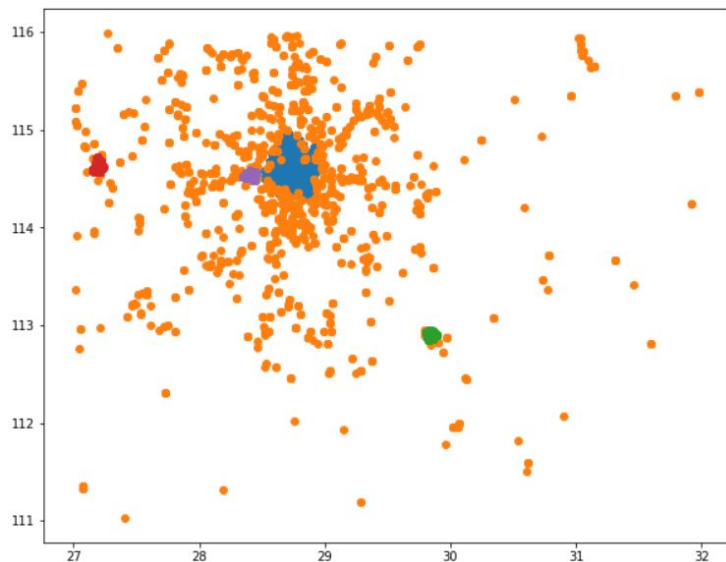
# Гипотезы и предположения

Предположение: Длительность неполадок коррелирует с географическим положением

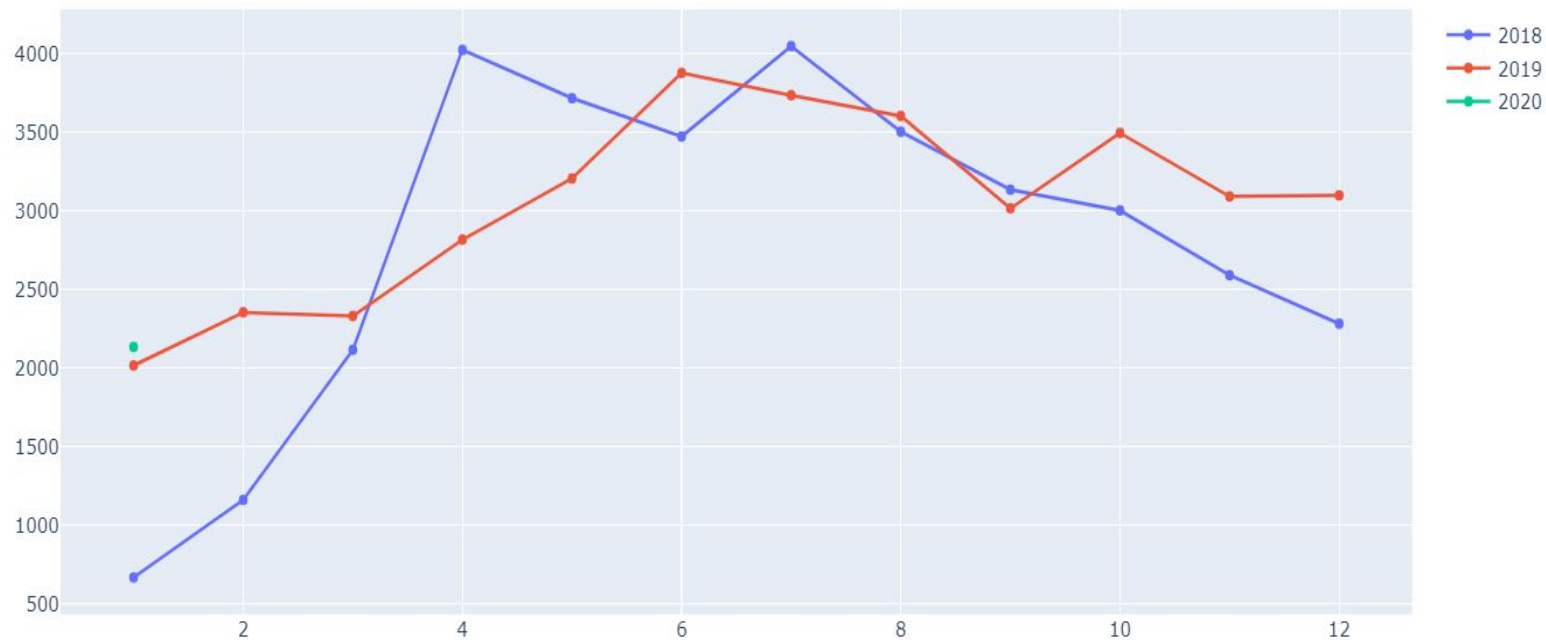


# Подход к кластеризации

- Использовать алгоритм DBSCAN
- Большие классы разбить вторично алгоритмом AffinityPropagation



# Формирование сезонных признаков



# Формирование сезонных признаков

1. year
2. month
3. day
4. weekday
5. hour



# Генерация дополнительных признаков

По каждому сектору

`mean_duration_time` - средняя длительность сбоя за предыдущий месяц

`min_duration_time` - минимальная длительность сбоя за предыдущий месяц

`max_duration_time` - максимальная длительность сбоя за предыдущий месяц

`mode_duration_time` - мода длительности сбоя за предыдущий месяц



# Алгоритмы ML

Мы пробовали:

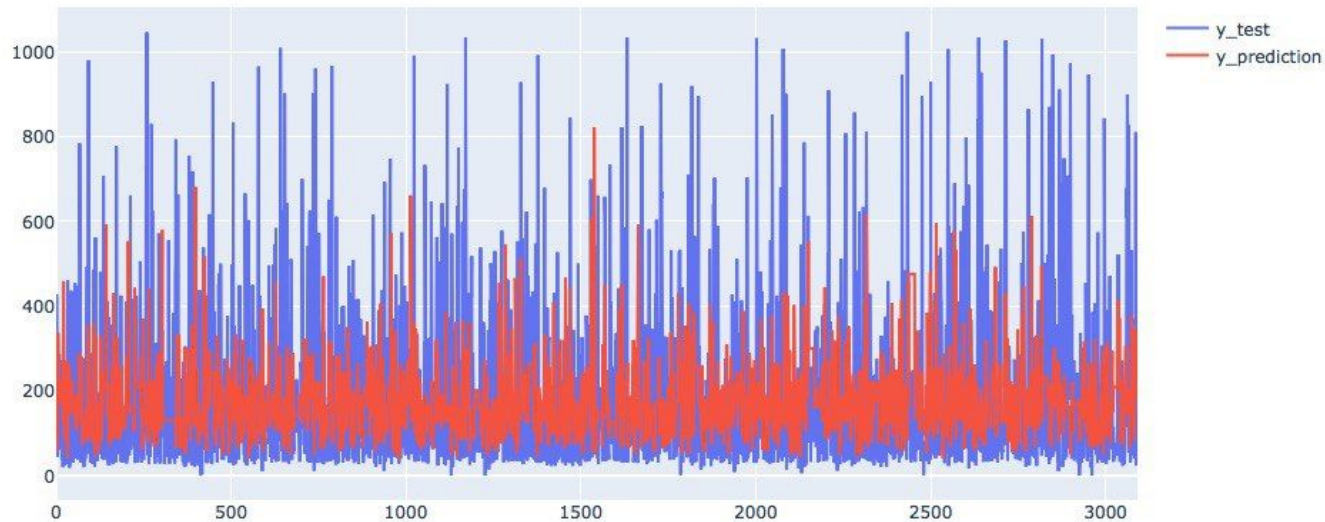
1. XGBoost
2. Случайный лес
3. Линейная регрессия





# Алгоритмы ML

Лучшую оценку показала линейная регрессия.



# Результат



Данные сильно зашумлены и требуют более глубокой обработки для улучшения прогнозирующей способности модели.



# Состав команды

Алан Тедеев [Analyst]

Чермен Валиев [Manager]

Станислав Теблоев [Tech]

