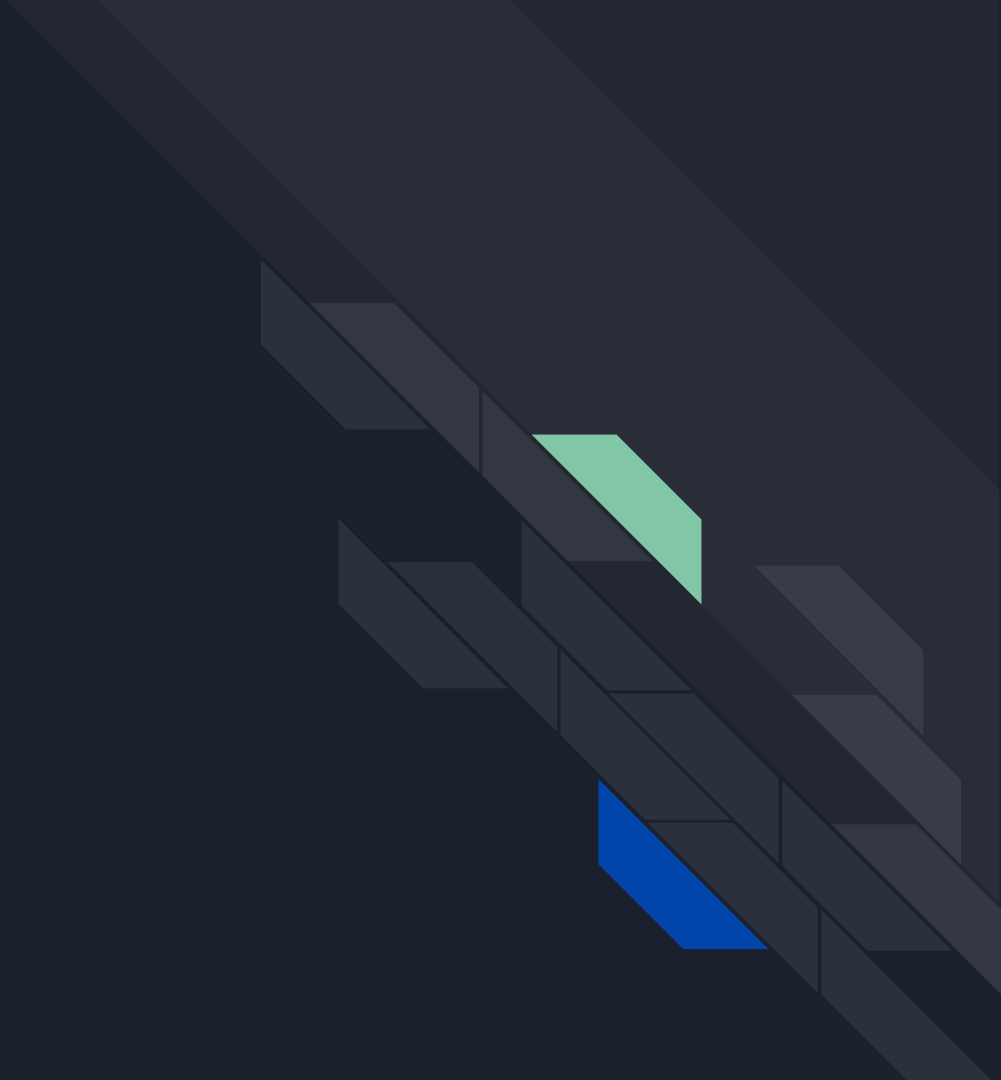
A decorative graphic on the left side of the slide consisting of two overlapping parallelograms. The front one is blue and the back one is a light green color. They are positioned diagonally, with the blue one partially covering the green one.

# Advanced Time Series Analysis

Prediction of Concurrent Radio Stream  
Listeners

by André Kern

# The Idea



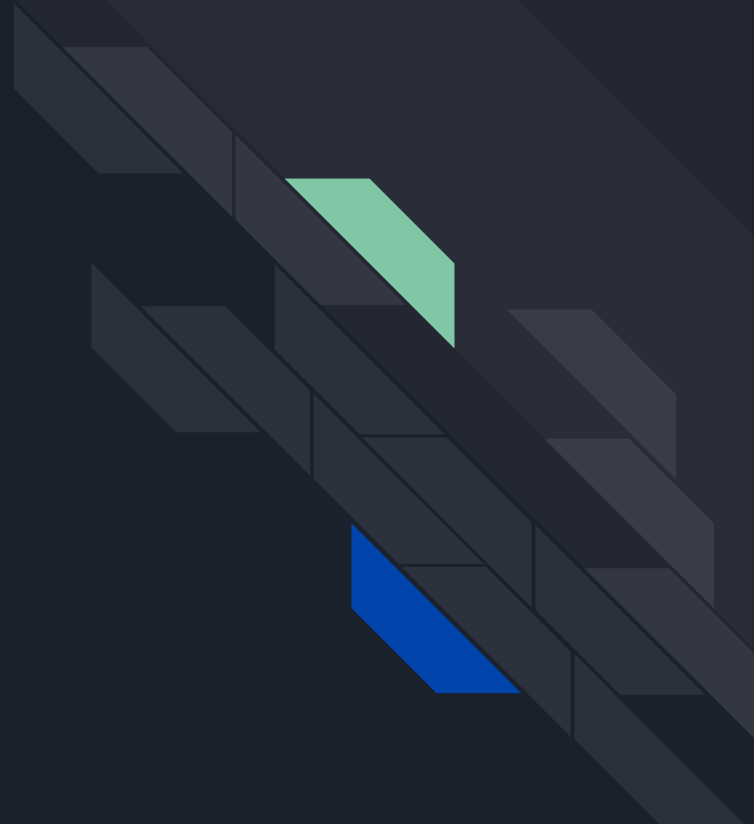


# The Idea

*How impacts the design of the musical program (internal factors) the listener counts?*

→ Investigating listener counts based on *external Factors* like time related features

# Dataset Description



# Dataset Description

Original Dataframe:

	ip	iphash	channelkey	channel	endtime	starttime
0	77.179.87.0	72e70ddcc78a9b6637fa06dbabfd2dd75600fce5a1b3d0f8991992bbd1042896	regc_pppcracqhx_cfhb	RADIO BOB - Tobias Sammet Rockshow	2024- 03-01 00:01:08	2024-03- 01 00:00:00
1	77.179.87.0	72e70ddcc78a9b6637fa06dbabfd2dd75600fce5a1b3d0f8991992bbd1042896	regc_ekm9wrc1s_s5lu	RADIO BOB - Wacken Nonstop	2024- 03-01 00:01:08	2024-03- 01 00:00:00
2	77.179.87.0	72e70ddcc78a9b6637fa06dbabfd2dd75600fce5a1b3d0f8991992bbd1042896	regc_lhwi0uadx6_j9g3	RADIO BOB - Hardrock	2024- 03-01 00:01:08	2024-03- 01 00:00:00
3	77.179.87.0	72e70ddcc78a9b6637fa06dbabfd2dd75600fce5a1b3d0f8991992bbd1042896	regc_v4jativbuc_kz1i	RADIO BOB - Best of Rock	2024- 03-01 00:01:09	2024-03- 01 00:00:00
4	77.179.87.0	72e70ddcc78a9b6637fa06dbabfd2dd75600fce5a1b3d0f8991992bbd1042896	regc_6zuinqoko_oskw	RADIO BOB - Christmas Rock	2024- 03-01 00:01:09	2024-03- 01 00:00:00



# Dataset Description

Transformed Data:

timestamp	year	month	week	hour	minute	second	day_week	day_month	daytime	concurrent_sessions	concurrent_sessions_7D
2024-03-08 00:00:12	2024	3	10	0	0	12	5	4	1	603	NaN
2024-03-08 00:00:13	2024	3	10	0	0	13	5	4	1	603	NaN
2024-03-08 00:00:14	2024	3	10	0	0	14	5	4	1	603	NaN
2024-03-08 00:00:15	2024	3	10	0	0	15	5	4	1	603	NaN
2024-03-08 00:00:16	2024	3	10	0	0	16	5	4	1	603	NaN
...	...	...	...	...	...	...	...	...	...	...	...
2024-04-30 23:59:55	2024	4	18	23	59	55	2	1	8	939	936.0
2024-04-30 23:59:56	2024	4	18	23	59	56	2	1	8	939	936.0
2024-04-30 23:59:57	2024	4	18	23	59	57	2	1	8	939	936.0
2024-04-30 23:59:58	2024	4	18	23	59	58	2	1	8	939	937.0
2024-04-30 23:59:59	2024	4	18	23	59	59	2	1	8	939	937.0

4665564 rows x 9 columns

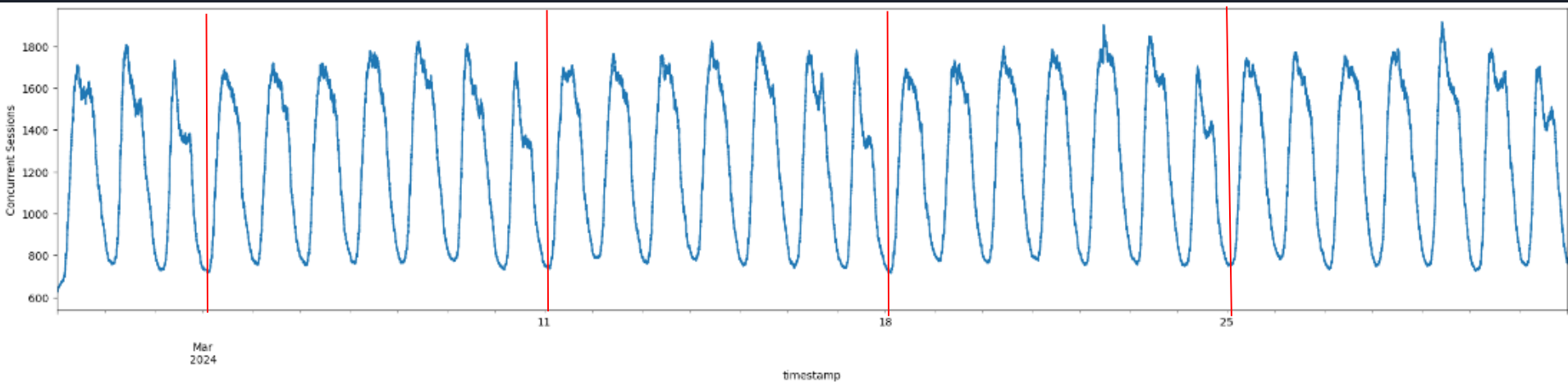


# Dataset Description

Transformed Data:

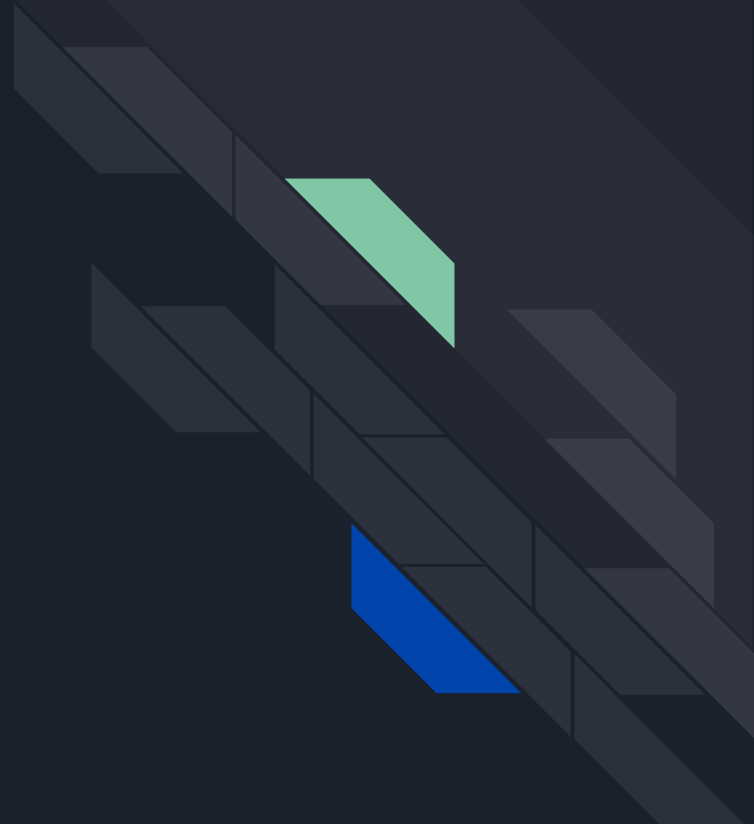
```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 4665564 entries, 2024-03-08 00:00:12 to 2024-04-30 23:59:59
Data columns (total 9 columns):
#   Column      Dtype
---  -
0   year        int64
1   month       int64
2   week        int64
3   hour        int64
4   minute      int64
5   second      int64
6   day_week    int64
7   day_month   int64
8   daytime     int64
dtypes: int64(9)
memory usage: 356.0 MB
```

# Dataset Description





# The Approach





# The Approach

## Methods

- XGBoost
- LightGBM



# The Approach

## Methods

- XGBoost
- LightGBM

## Features

- Time Related Features
  - Second, Minute, Hour, Daytime, Weekday, Week, Month, Year



# The Approach

## Methods

- XGBoost
- LightGBM

## Features

- Time Related Features
  - Second, Minute, Hour, Daytime, Weekday, Week, Month, Year

## Target

- Concurrent Sessions



# The Approach

## Methods

- XGBoost
- LightGBM

## Features

- Time Related Features
  - Second, Minute, Hour, Daytime, Weekday, Week, Month, Year

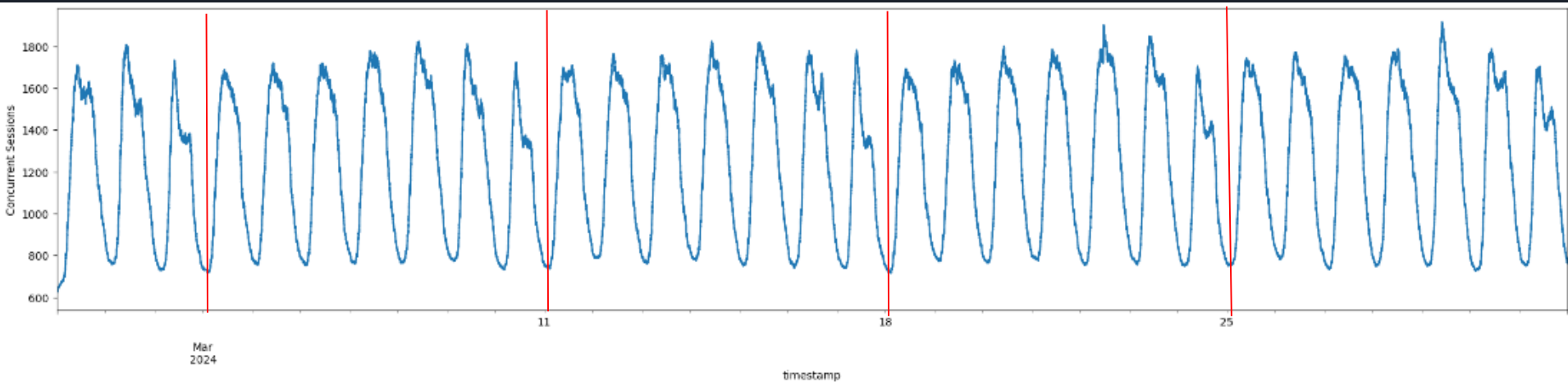
## Target

- Concurrent Sessions

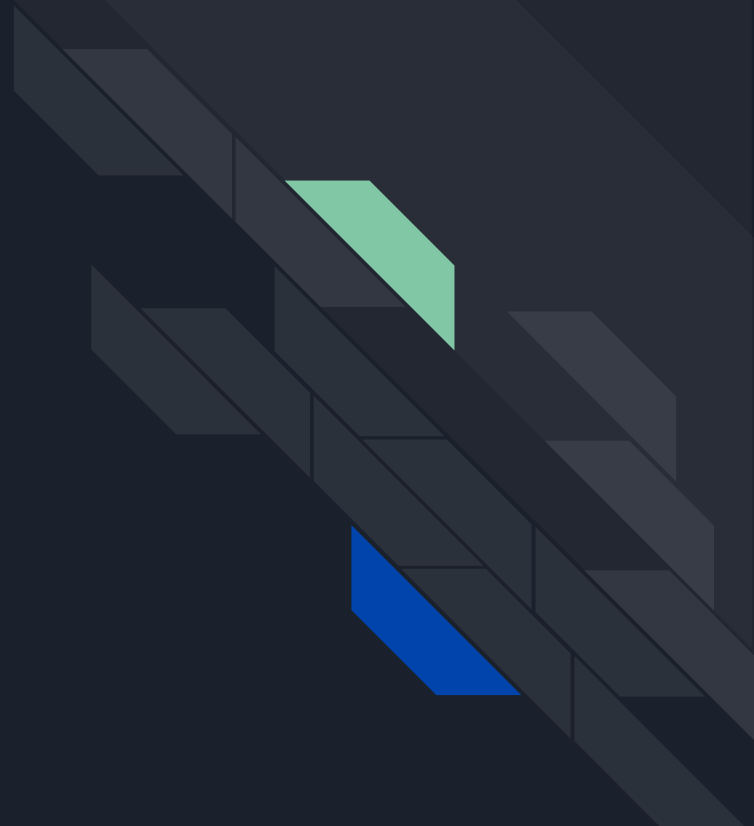
## Baseline

- 7 Days Shifted Concurrent Sessions

# Dataset Description



# Evaluation





# Evaluation

	modelname	target	features	mse	mae	r2	explained_variance
0							
1	XGBoost_default	concurrent_sessions	[year, 'month', 'week', 'hour', 'minute', 'second', 'day_week', 'day_month', 'daytime']	3206.561853	38.182828	0.974152	0.974425
2	LightGBM_default	concurrent_sessions	[year, 'month', 'week', 'hour', 'minute', 'second', 'day_week', 'day_month', 'daytime']	2816.081302	34.944549	0.977300	0.977752
3							
4							





# Model Training - XGBoost

```
xgb_model_opt = BayesSearchCV(  
    XGBRegressor(device='gpu'),  
    {  
        'learning_rate': (10e-6, 1.0, 'log-uniform'),  
        'max_depth': Integer(0, 50, 'uniform'),  
        'n_estimators' : (10, 1000, 'log-uniform'),  
    },  
    n_iter=10,  
    cv=3  
)
```

# Model Training - XGBoost

```
xgb_model_opt = BayesSearchCV(  
    XGBRegressor(device='gpu'),  
    {  
        'learning_rate': (10e-6, 1.0, 'log-uniform'),  
        'max_depth': Integer(0, 50, 'uniform'),  
        'n_estimators' : (10, 1000, 'log-uniform'),  
    },  
    n_iter=10,  
    cv=3  
)
```

▼ XGBRegressor

```
XGBRegressor(base_score=None, booster=None, callbacks=None,  
             colsample_bylevel=None, colsample_bynode=None,  
             colsample_bytree=None, device=None, early_stopping_rounds=None,  
             enable_categorical=False, eval_metric=None, feature_types=None,  
             gamma=None, grow_policy=None, importance_type=None,  
             interaction_constraints=None, learning_rate=0.21665653338912771,  
             max_bin=None, max_cat_threshold=None, max_cat_to_onehot=None,  
             max_delta_step=None, max_depth=26, max_leaves=None,  
             min_child_weight=None, missing=nan, monotone_constraints=None,  
             multi_strategy=None, n_estimators=31, n_jobs=None,  
             num_parallel_tree=None, random_state=None, ...)
```



# Model Training - LightGBM

```
lgbm_model_opt = BayesSearchCV(  
    LGBMRegressor(),  
    {  
        'learning_rate': (10e-6, 1.0, 'log-uniform'),  
        'max_depth': Integer(-1, 50, 'uniform'),  
        'n_estimators' : (10, 1000, 'log-uniform'),  
        'num_leaves': Integer(2, 256, 'uniform'),  
    },  
    n_iter=10,  
    cv=3  
)
```



# Model Training - LightGBM

```
lgbm_model_opt = BayesSearchCV(  
    LGBMRegressor(),  
    {  
        'learning_rate': (10e-6, 1.0, 'log-uniform'),  
        'max_depth': Integer(-1, 50, 'uniform'),  
        'n_estimators' : (10, 1000, 'log-uniform'),  
        'num_leaves': Integer(2, 256, 'uniform'),  
    },  
    n_iter=10,  
    cv=3  
)
```

LGBMRegressor

LGBMRegressor(learning\_rate=0.29271345926800346, max\_depth=25, n\_estimators=10,  
num\_leaves=45, seed=42)



# Evaluation

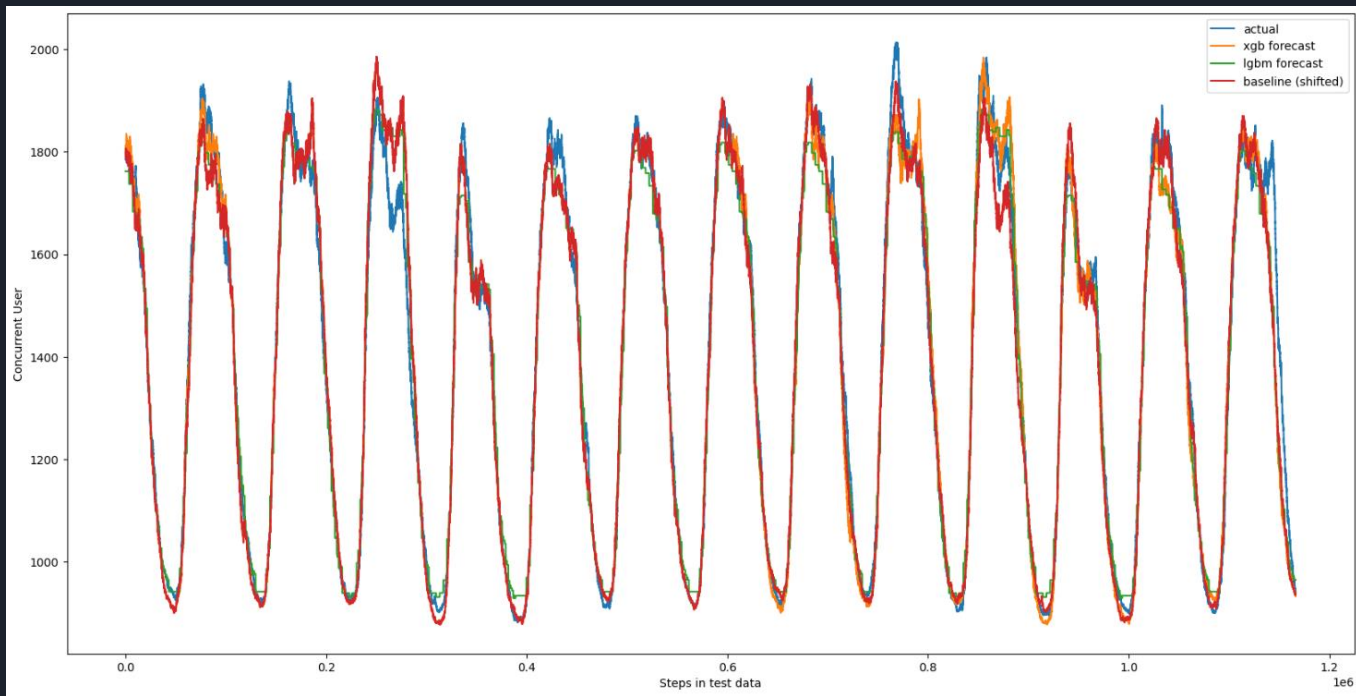
	modelname	target	features	mse	mae	r2	explained_variance
0							
1	XGBoost_default	concurrent_sessions	[year, 'month', 'week', 'hour', 'minute', 'second', 'day_week', 'day_month', 'daytime]	3206.561853	38.182828	0.974152	0.974425
2	LightGBM_default	concurrent_sessions	[year, 'month', 'week', 'hour', 'minute', 'second', 'day_week', 'day_month', 'daytime]	2816.081302	34.944549	0.977300	0.977752
3	XGBoost_optimized	concurrent_sessions	[year, 'month', 'week', 'hour', 'minute', 'second', 'day_week', 'day_month', 'daytime]	3226.341378	37.913312	0.973992	0.974447
4	LightGBM_optimized	concurrent_sessions	[year, 'month', 'week', 'hour', 'minute', 'second', 'day_week', 'day_month', 'daytime]	3114.260411	39.674587	0.974896	0.975670



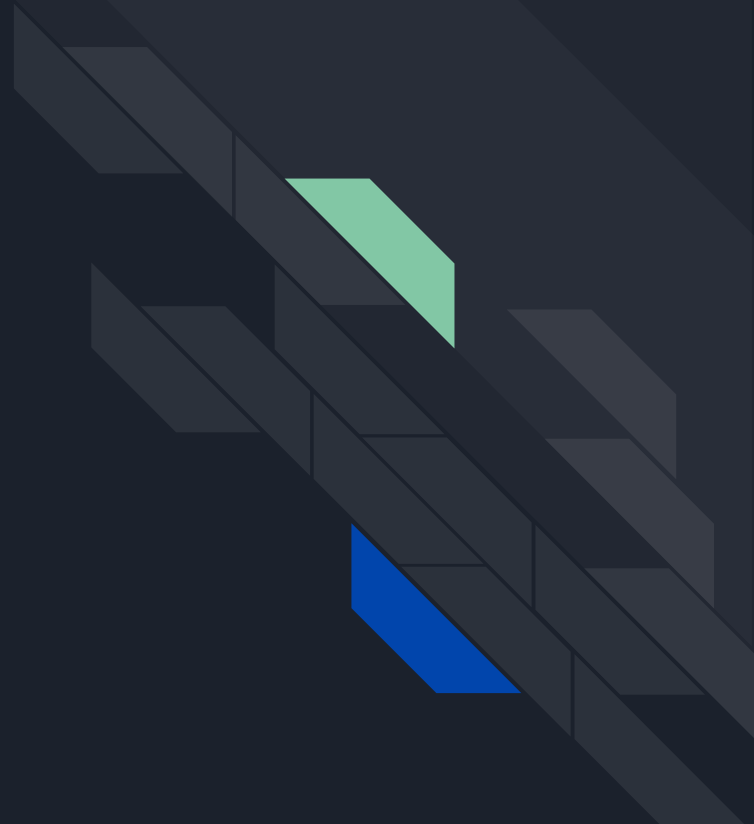
# Evaluation

	modelName	target	features	mse	mae	r2	explained_variance
0	baseline_model	concurrent_sessions	concurrent_sessions_7D	3099.132959	37.412436	0.975018	0.975988
1	XGBoost_default	concurrent_sessions	[year, 'month', 'week', 'hour', 'minute', 'second', 'day_week', 'day_month', 'daytime']	3206.561853	38.182828	0.974152	0.974425
2	LightGBM_default	concurrent_sessions	[year, 'month', 'week', 'hour', 'minute', 'second', 'day_week', 'day_month', 'daytime']	2816.081302	34.944549	0.977300	0.977752
3	XGBoost_optimized	concurrent_sessions	[year, 'month', 'week', 'hour', 'minute', 'second', 'day_week', 'day_month', 'daytime']	3226.341378	37.913312	0.973992	0.974447
4	LightGBM_optimized	concurrent_sessions	[year, 'month', 'week', 'hour', 'minute', 'second', 'day_week', 'day_month', 'daytime']	3114.260411	39.674587	0.974896	0.975670

# Evaluation



# Problems Encountered



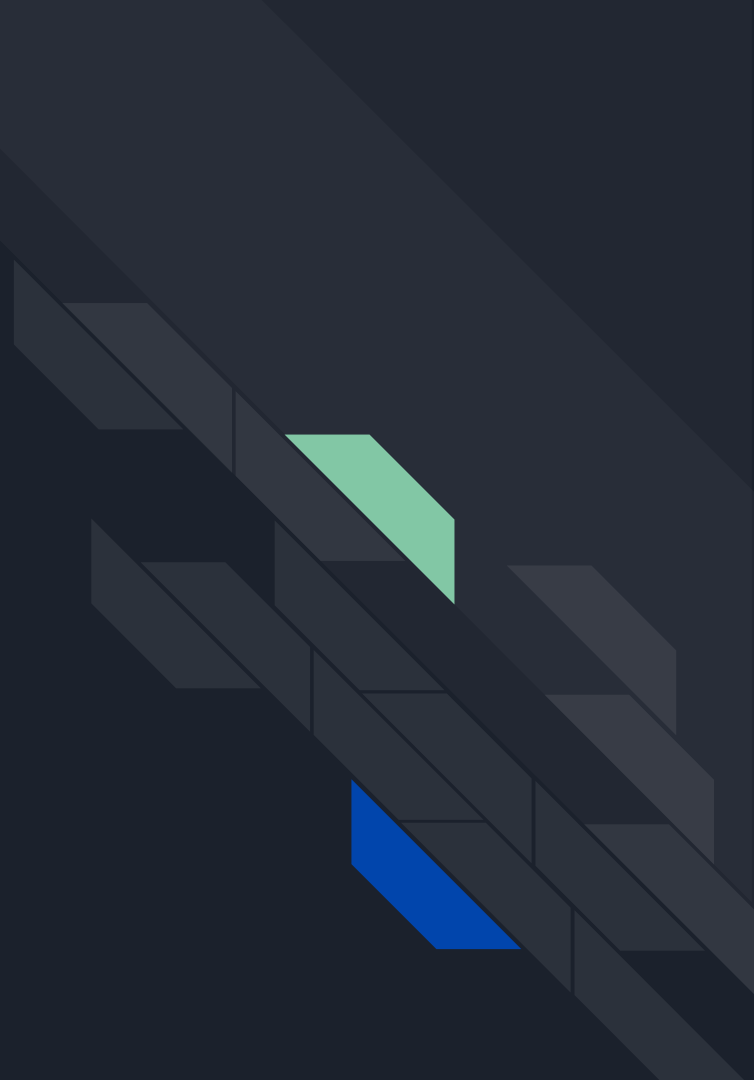




# Problems Encountered

1. Large Dataset
2. Too Many Timestamps
3. Inapplicability of Certain Algorithms

Ask me your questions!



Thank you for your  
attention!

