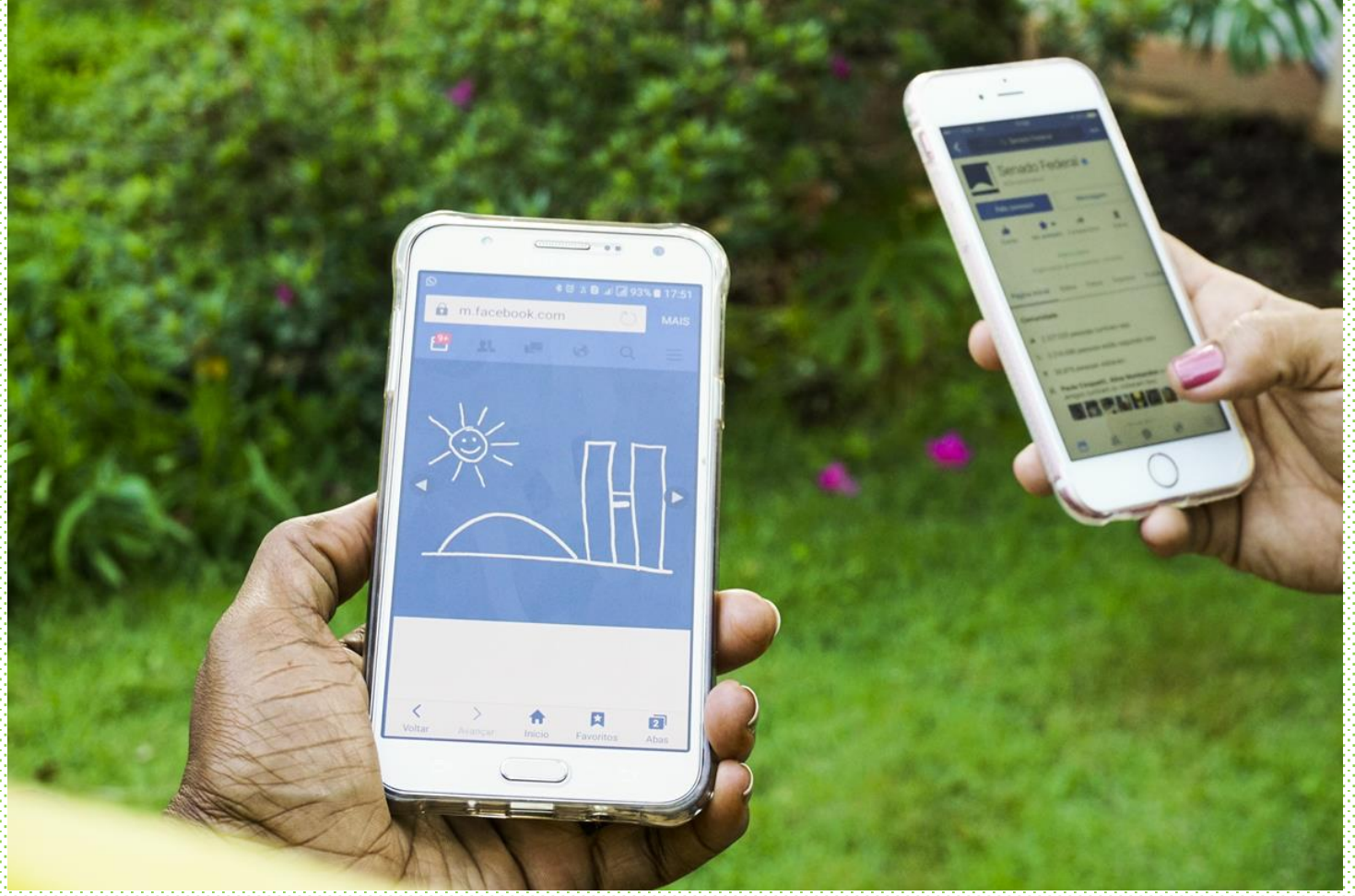


PREDICTING HUMAN ACTIVITIES USING SMARTPHONES

(Data Science Capstone)

PRESENTATION FLOW

- Explore the research question
- Discuss the Dataset Chosen
- Identify the Purpose of the Study
- Exploratory analysis of Data
- Classification Models performance comparison
- Concluding Remarks

WHAT DOES YOUR SMARTPHONE KNOW ABOUT YOU?

ABOUT DATASET

The Human Activity Recognition database was built from the recordings of 30 study participants performing activities of daily living (ADL) while carrying a waist-mounted smartphone with embedded inertial sensors.

Attribute information :

- Triaxial acceleration from the accelerometer (total acceleration) and the estimated body acceleration.
- Triaxial Angular velocity from the gyroscope.
- A 561-feature vector with time and frequency domain variables.
- Its activity label.
- An identifier of the subject who carried out the experiment.

THE PURPOSE OF THE STUDY

To classify activities into one of the six activities performed.

Tracking Human Activities :

- WALKING
- WALKING_UPSTAIRS
- WALKING_DOWNSTAIRS
- SITTING
- STANDING
- LAYING

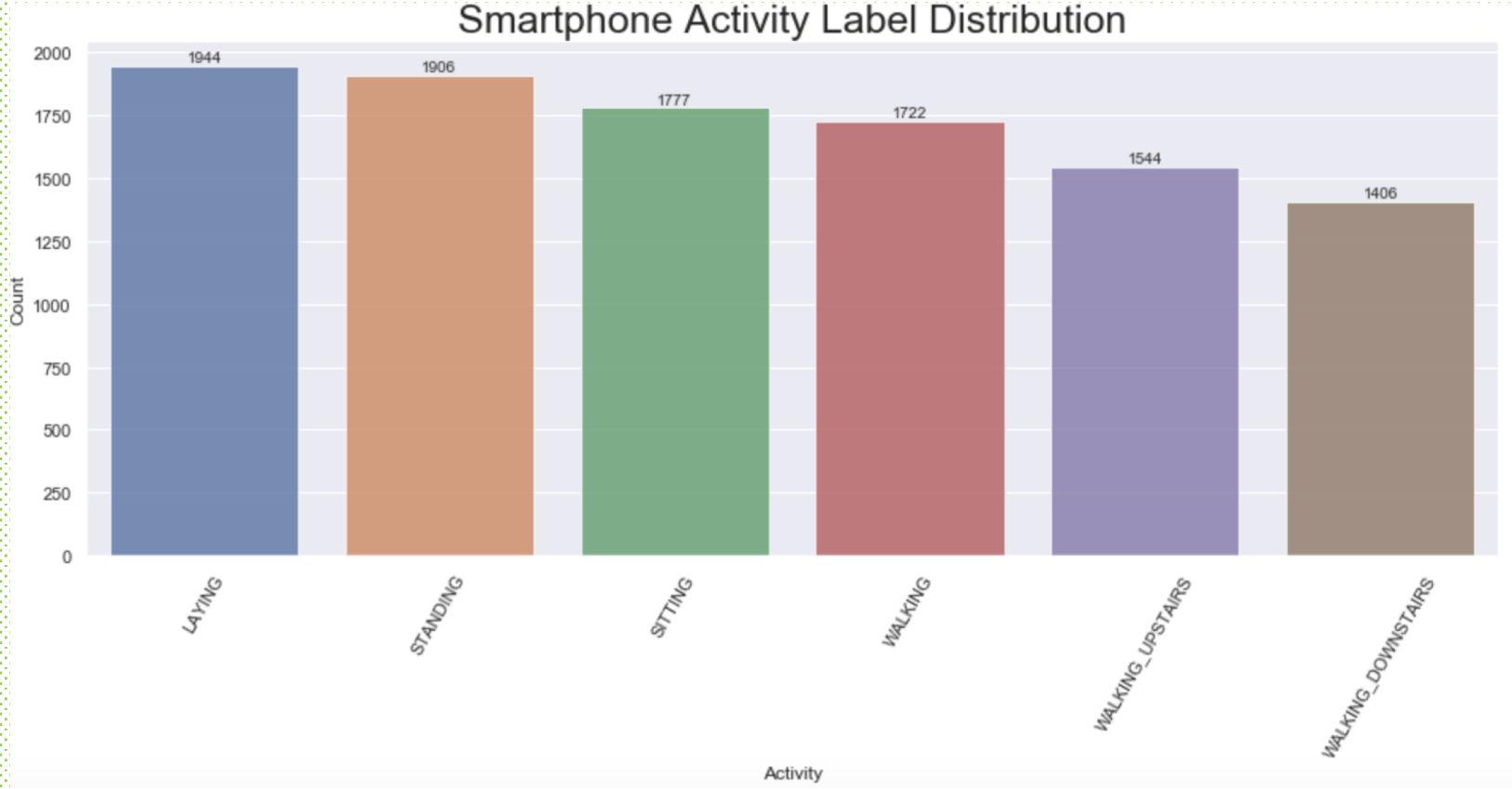
EXPLORATORY ANALYSIS OF DATA

- Mainly there are 'acceleration' and 'gyroscope' features. A few 'gravity' features as well.
- Total 563 features
- Except 'Activity' and 'subject' features there is only numerical data. Fortunately there are no missing values.
- The features seem to have a main name and some information on how they have been computed attached

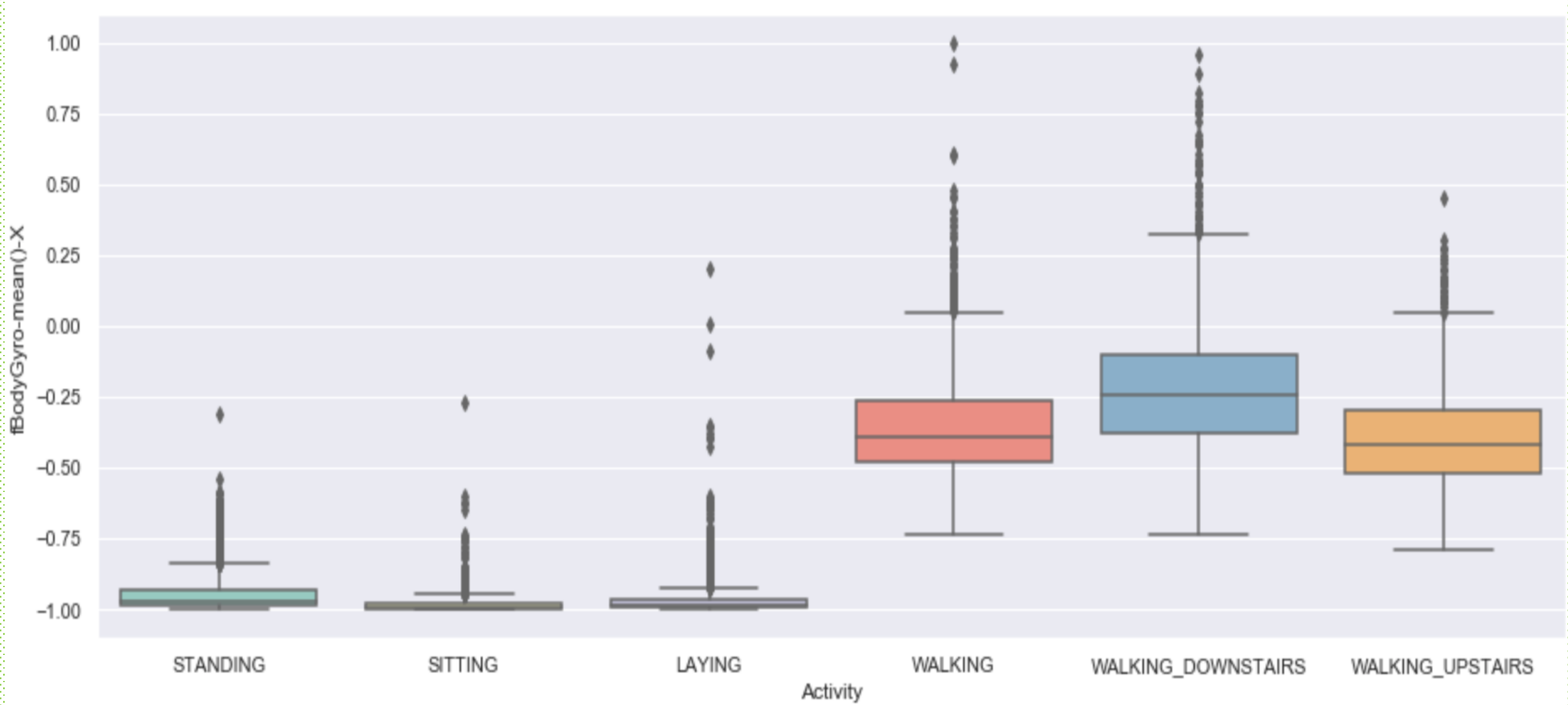
Data: https://www.kaggle.com/morrisb/what-does-your-smartphone-know-about-you/data

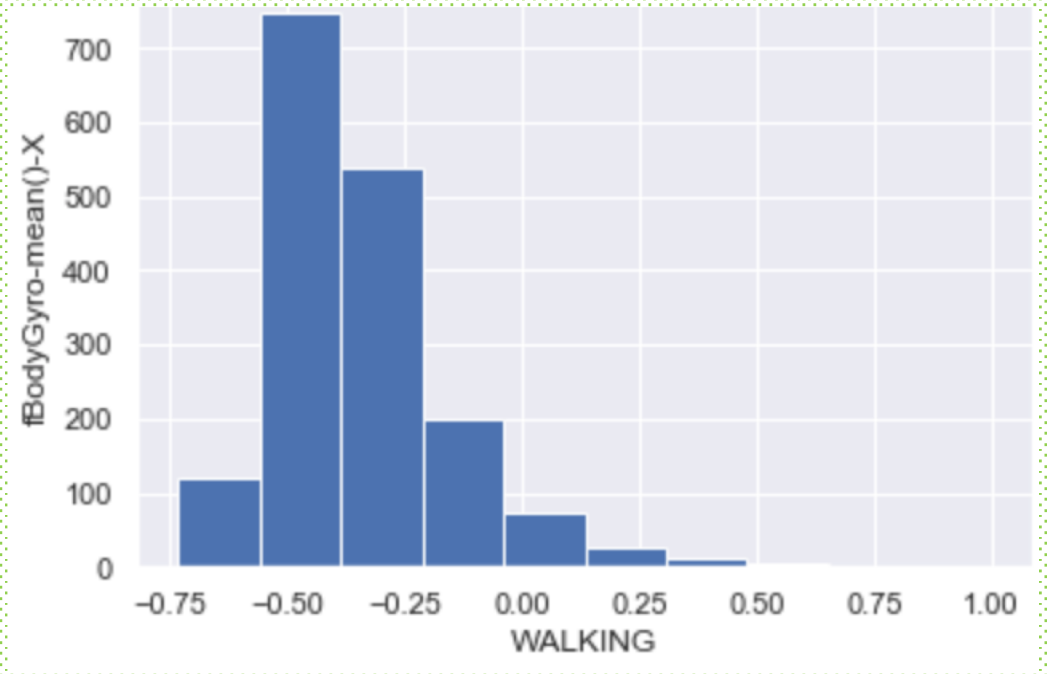
	count
fBodyAcc	79
fBodyGyro	79
fBodyAccJerk	79
tGravityAcc	40
tBodyAcc	40
tBodyGyroJerk	40
tBodyGyro	40
tBodyAccJerk	40
tBodyAccMag	13
tGravityAccMag	13
tBodyAccJerkMag	13
tBodyGyroMag	13
tBodyGyroJerkMag	13
fBodyAccMag	13
fBodyBodyAccJerkMag	13
fBodyBodyGyroMag	13
fBodyBodyGyroJerkMag	13
angle	7
subject	1
Activity	1

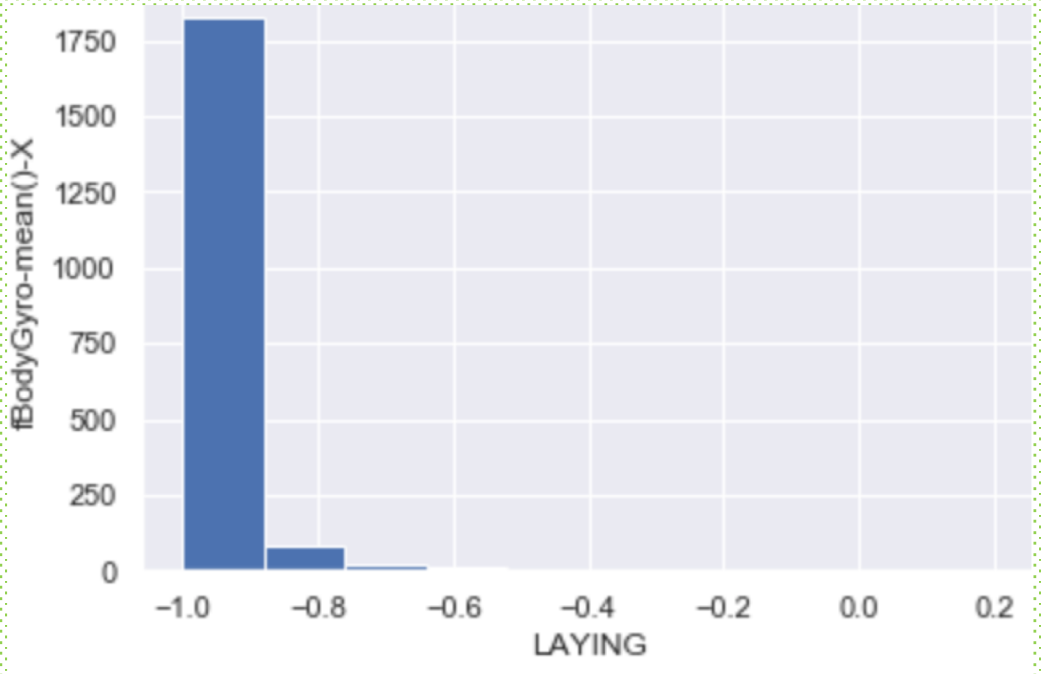
DATA ANALYSIS

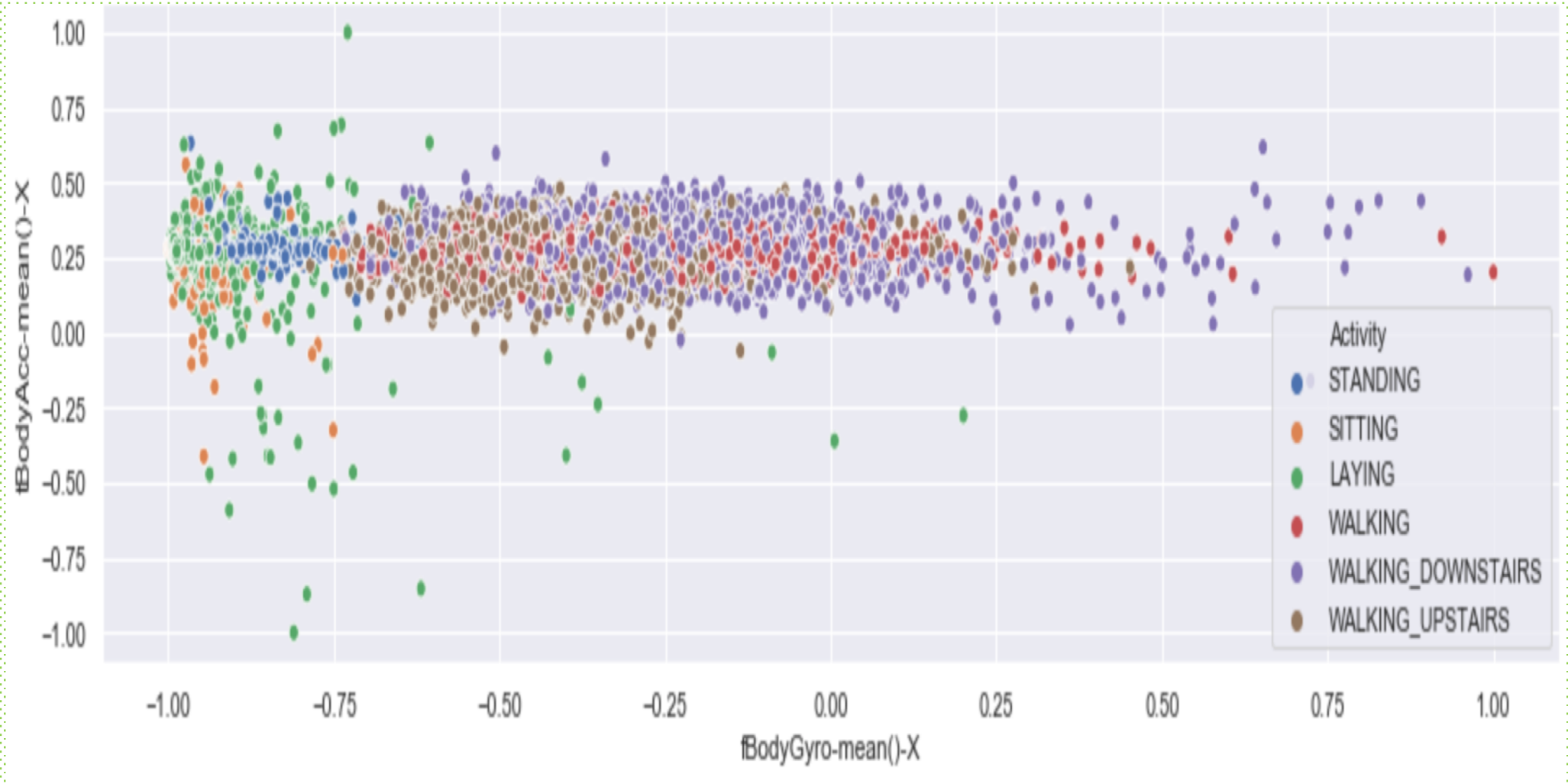


Disregarding the possibility of flawed data, the participants seem to **walk roughly 10% faster downwards.**

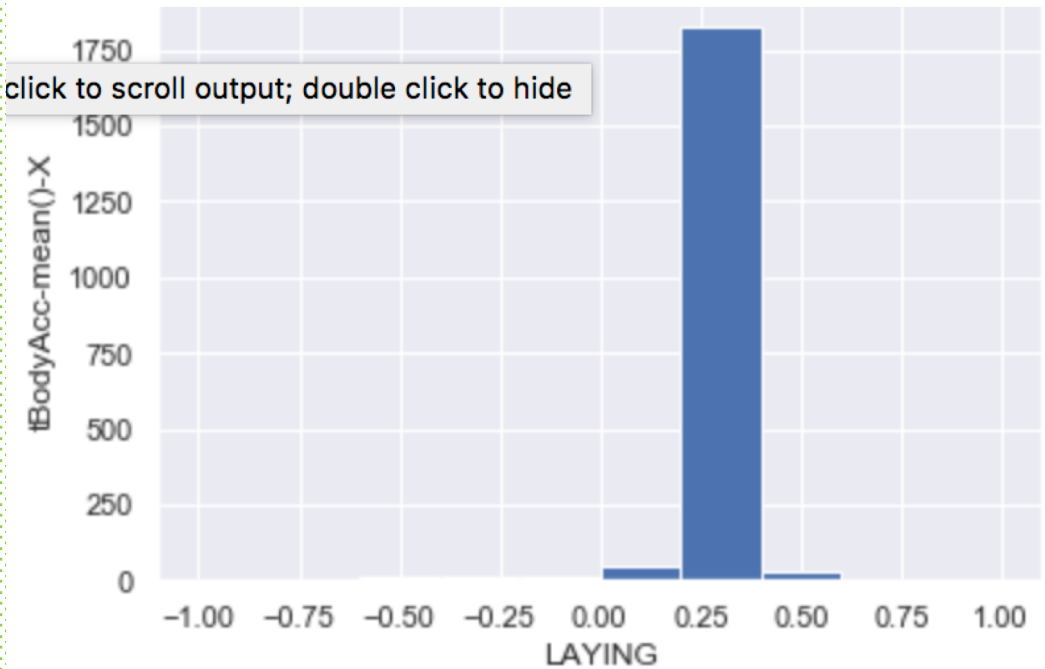

1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
fBodyGyro-mean()-X
STANDING
SITTING
LAYING
WALKING
WALKING_DOWNSTAIRS
WALKING_UPSTAIRS
Activity

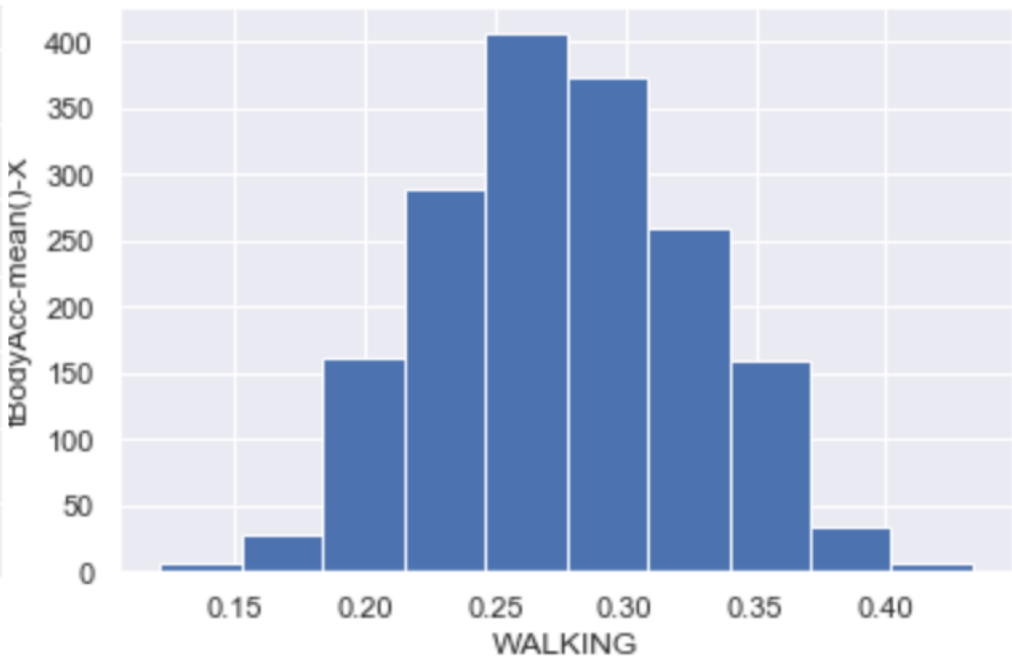

fBodyGyro-mean()-X
700
600
500
400
300
200
100
0
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00
WALKING


fBodyGyro-mean()-X
1750
1500
1250
1000
750
500
250
0
−1.0
−0.8
−0.6
−0.4
−0.2
0.0
0.2
LAYING


1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
fBodyAcc-mean()-X
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00
fBodyGyro-mean()-X
Activity
STANDING
SITTING
LAYING
WALKING
WALKING_DOWNSTAIRS
WALKING_UPSTAIRS

click to scroll output; double click to hide





CLASSIFICATION MODELS PERFORMANCE COMPARISON

- Before PCA

Model	Score
GradientBoostingClassifier	0.984466
LogisticRegression	0.983689
XGBClassifier	0.983301
RandomForestClassifier	0.969709
KNeighborsClassifier	0.961942
SVC	0.958835
DecisionTreeClassifier	0.931650
GaussianNB	0.730874

- After PCA

Model	Score
LogisticRegression	0.953398
KNeighborsClassifier	0.946019
GradientBoostingClassifier	0.940583
XGBClassifier	0.927767
SVC	0.919223
RandomForestClassifier	0.894369
DecisionTreeClassifier	0.838835
GaussianNB	0.833398

CONCLUSIONS AFTER PCA

- Data analysis shows that PCA is not helping to improve model accuracy as there must be some loss of information so we are not considering PCA for improving model performance
- The best performing model is Logistic Regression with improved performance after tuning and accuracy of 98%
- Vanilla gradient boosting without tuning performed the best, with tuning it didn't help much
- Logistic Regression is way faster than Gradient boosting

GRADIENT BOOSTING CLASSIFIER - RESULT

	precision	recall	f1-score	support
LAYING	1.00	1.00	1.00	502
SITTING	0.95	0.95	0.95	443
STANDING	0.96	0.95	0.95	457
WALKING	0.99	0.99	0.99	446
WALKING_UPSTAIRS	0.99	0.97	0.98	343
WALKING_DOWNSTAIRS	0.97	0.99	0.98	384
avg / total	0.98	0.98	0.98	2575

click to scroll output; double click to hide

	LAYING	SITTING	STANDING	WALKING	WALKING_UPSTAIRS	WALKING_DOWNSTAIRS
LAYING	502	0	0	0	0	0
SITTING	0	422	20	0	0	1
STANDING	0	21	434	0	0	2
WALKING	0	0	0	442	2	2
WALKING_UPSTAIRS	0	0	0	4	334	5
WALKING_DOWNSTAIRS	0	0	0	1	3	380

LOGISTIC REGRESSION - RESULT

	precision	recall	f1-score	support
LAYING	1.00	1.00	1.00	502
SITTING	0.95	0.95	0.95	443
STANDING	0.95	0.96	0.96	457
WALKING	1.00	1.00	1.00	446
WALKING_UPSTAIRS	1.00	1.00	1.00	343
WALKING_DOWNSTAIRS	1.00	1.00	1.00	384
avg / total	0.98	0.98	0.98	2575

	LAYING	SITTING	STANDING	WALKING	WALKING_UPSTAIRS	WALKING_DOWNSTAIRS
LAYING	502	0	0	0	0	0
SITTING	0	422	21	0	0	0
STANDING	0	20	437	0	0	0
WALKING	0	0	0	446	0	0
WALKING_UPSTAIRS	0	0	0	0	343	0
WALKING_DOWNSTAIRS	0	0	0	1	0	383

CONCLUDING REMARKS

- The smartphone has enough data to determine what its user is doing (**98%**: 6 activities)
- Logistic Regression is giving more accurate and improved performance compared to other classification models also faster compared to gradient boost
- This data can be used by many healthcare and fitness apps which can provide more insights
- As the dataset only provide a small amount of data the variance of the results can be high


Thank
you!!