SENSING TRANSPORT MODES WITH PHONE SENSORS

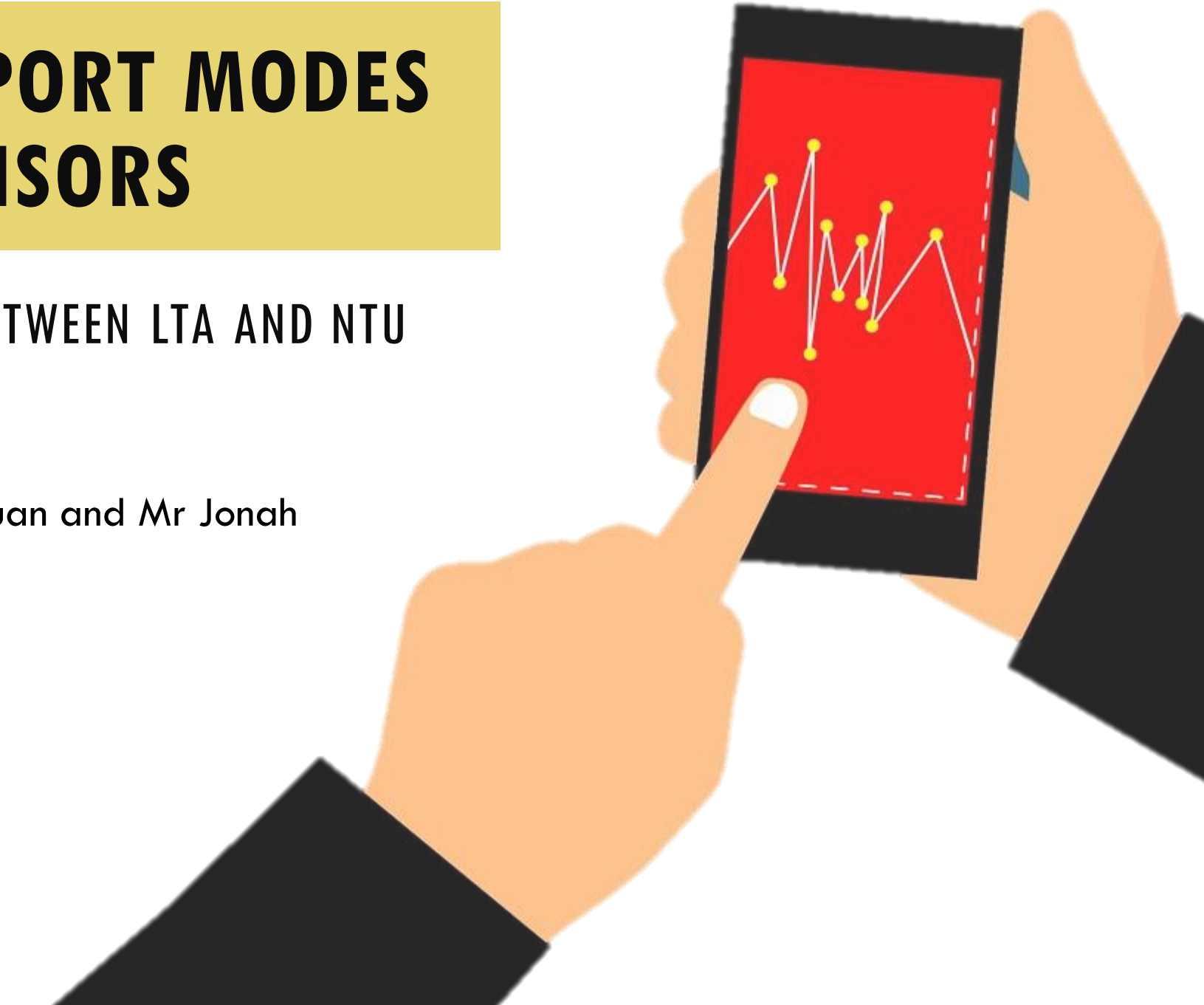
A COLLABORATIVE PROJECT BETWEEN LTA AND NTU

Presented by Voon Hao (Travis)

Guided by Ms Rachel, Mr Guangquan and Mr Jonah

28 Feb 2020

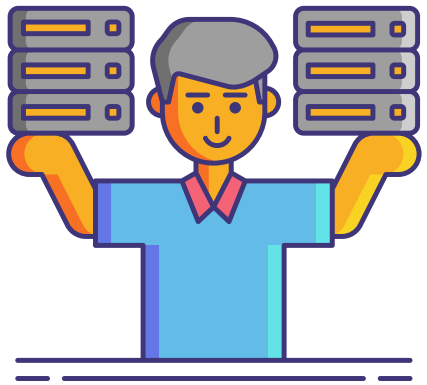
Land Transport Authority



BACKGROUND

The Mobility Sensing Project aims to identify travel patterns from GPS data and other multi-sensory data collected in smartphones to provide personalised travel information during service disruption and support the goals of LTMP 2040.

THE ROADMAP



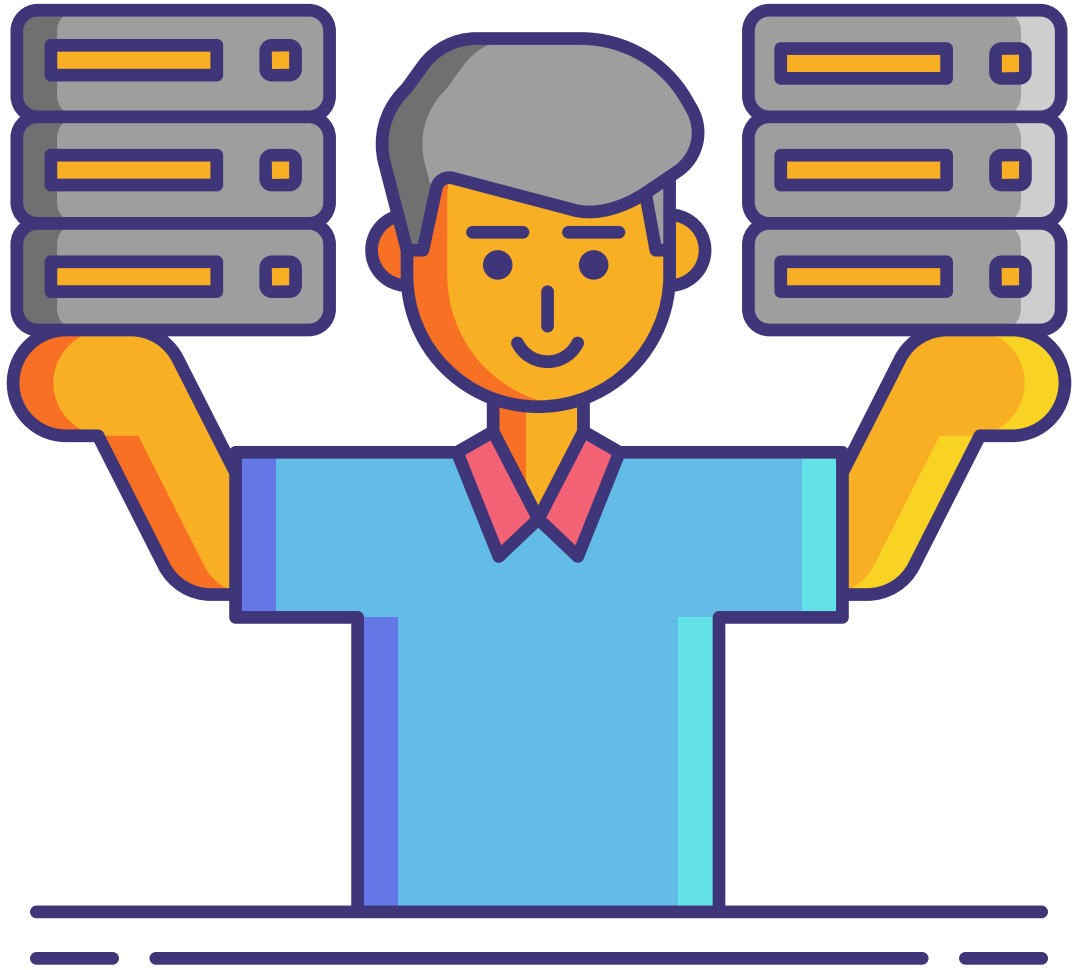
Data Collection

Data Preprocessing

Feature Selection

Model Training

PART 1
DATA COLLECTION



DATA COLLECTED AT A GLANCE

10 hours TOTAL DURATION	**400** DATA POINTS/S
~15,000,000 DATA POINTS PER SENSOR PER AXIS	
2 SENSORS	**3** AXIS

PART 2
DATA PREPROCESSING

AN OVERVIEW

1. Data Cleaning

2. Apply Fourier Transform for frequency-domain analysis

3. Extract features in the time- and frequency domains

DATA COLLECTION > DATA PREPROCESSING > FEATURE SELECTION > MODEL TRAINING

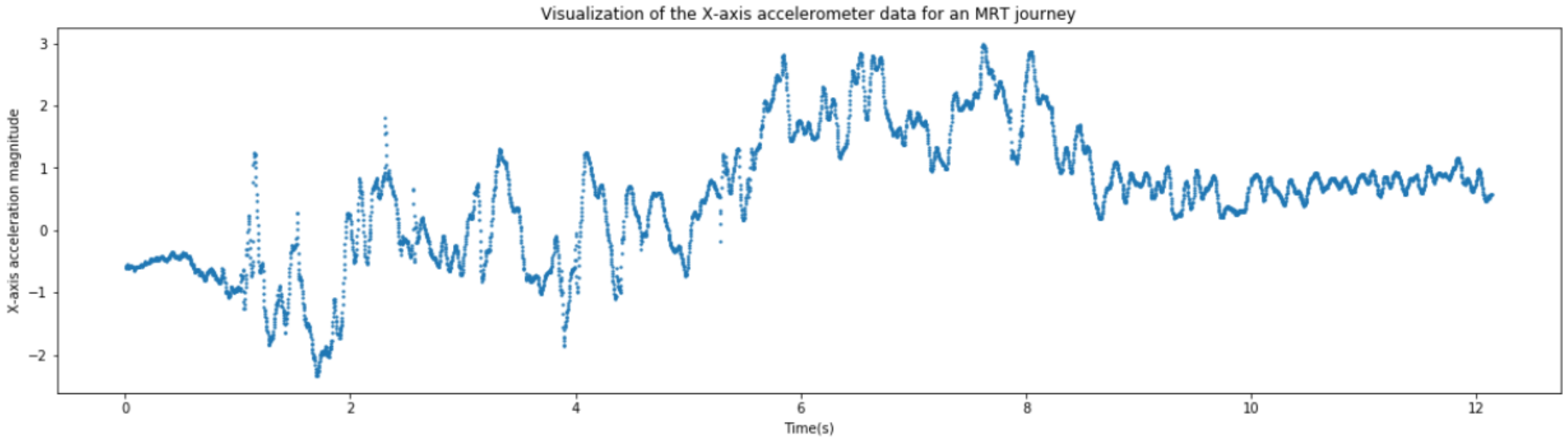
1. Data Cleaning
2. Fourier Transform
3. Extract summary statistics

1. Data Cleaning | 2. Fourier Transform | 3. Extract summary statistics

- Apply a sliding window of 2s with 50% overlap.
- Filtering data

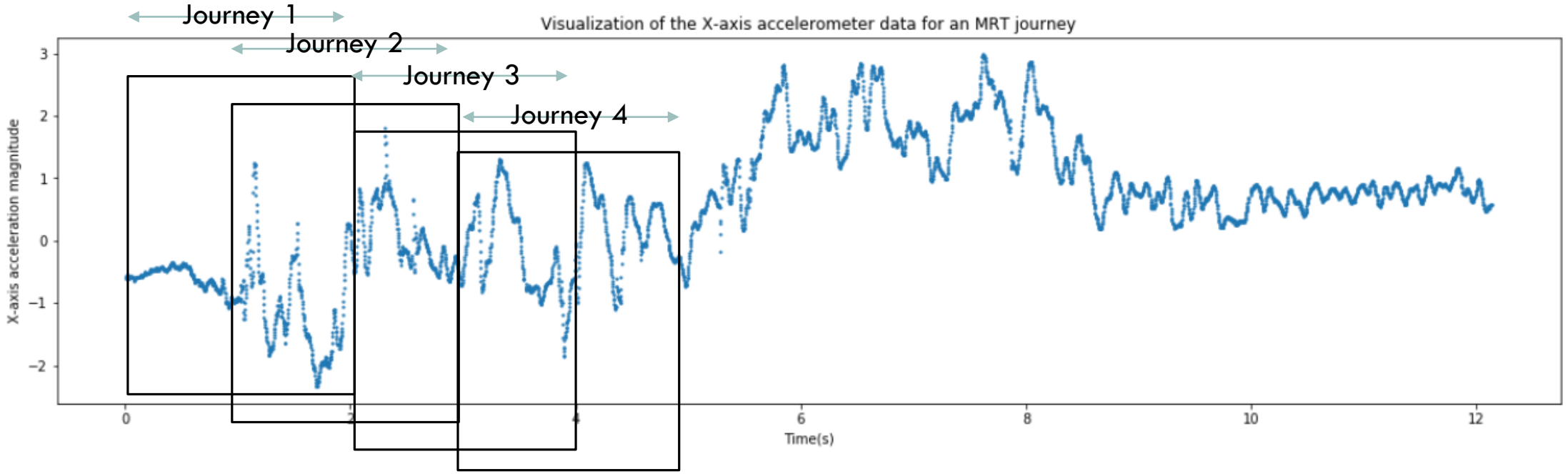
1. Data Cleaning | 2. Fourier Transform | 3. Extract summary statistics

- **Apply a sliding window of 2s with 50% overlap.**
- Filtering data



1. Data Cleaning | 2. Fourier Transform | 3. Extract summary statistics

- **Apply a sliding window of 2s with 50% overlap.**
- Filtering data



1. Data Cleaning | 2. Fourier Transform | 3. Extract summary statistics

- Apply a sliding window of 2s with 50% overlap.
- **Filtering data**

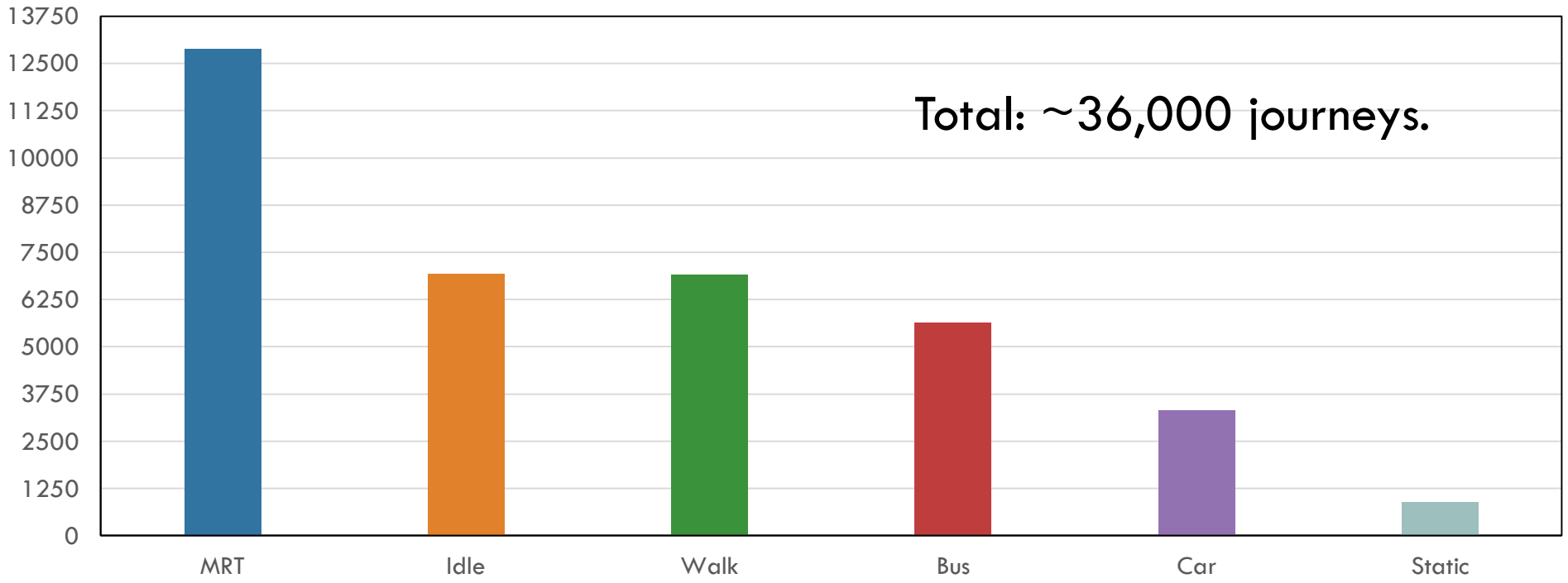
Rationale:

"Acceleration signals registered by smartphones are the sum of two components, a gravity component and a motion component… We may wish to subtract gravity to fully isolate body acceleration signals."

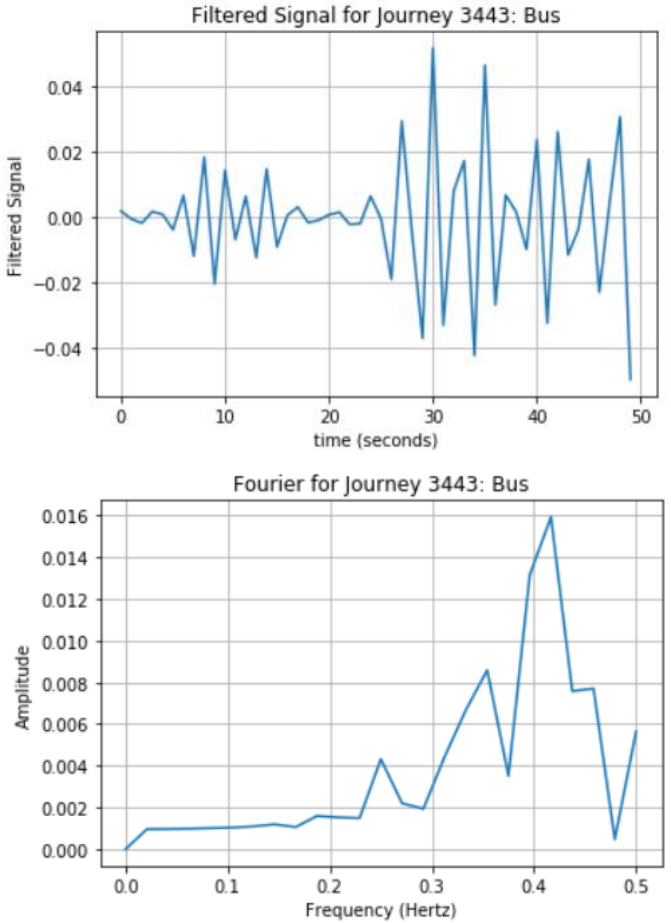
Physical activity recognition by smartphones, a survey (Morales, 2017)

1. Data Cleaning | 2. Fourier Transform | 3. Extract summary statistics

Number of 'journeys' for each transport mode

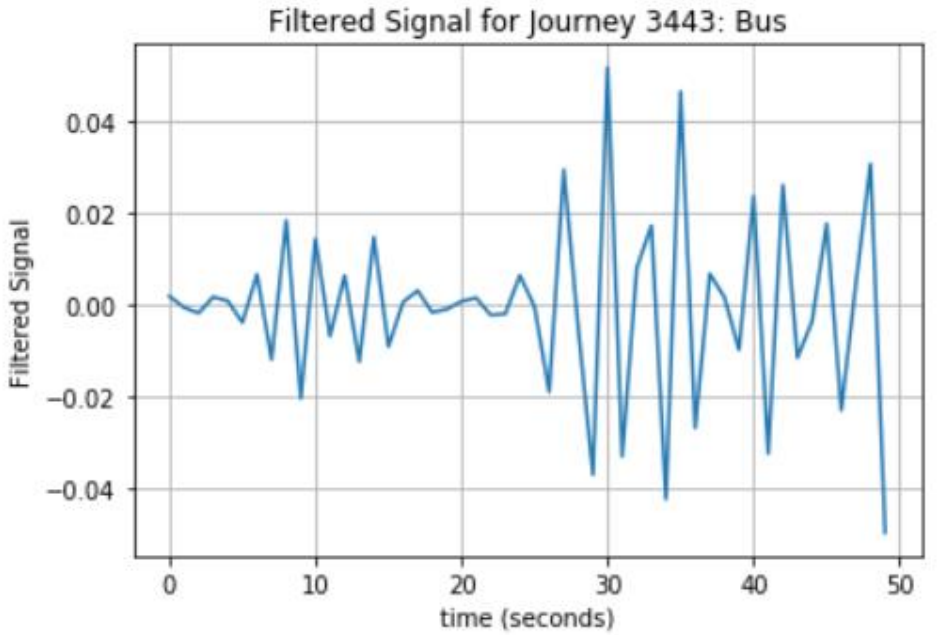


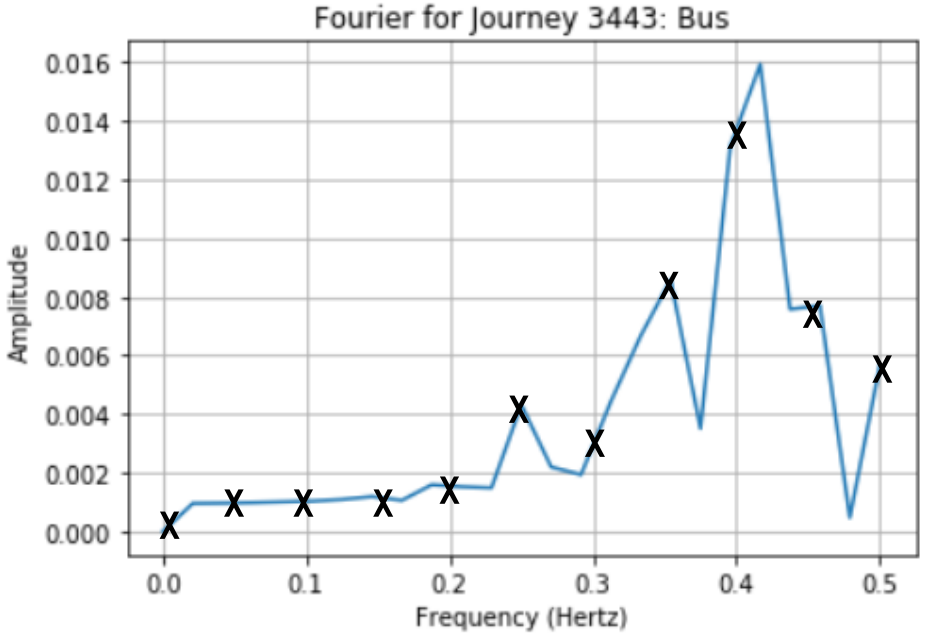
We can decompose a complicated signal into its components using **Fourier Transform.**



The **magnitudes** of Fourier Transform at each frequency are used as features.

For a vigorous mathematical understanding, please refer to Introduction to Fourier Transform.





1. Data Cleaning | 2. Fourier Transform | **3. Extract features**

Time-domain features were extracted.

No	Summary Statistic	Mathematical Intuition	Possible Usefulness
1	Mean	Average amplitude of acceleration for a journey	If a user is on an MRT, the magnitude of acceleration is higher than, say, if a bus.
2	Median		
3	RMS		
4	Skew	The shape of a probability distribution	-
5	Kurtosis		

Towards Clustering of Mobile and Smartwatch Accelerometer Data for Physical Activity Recognition (Dobbins, 2018)
Human Activity Recognition Using Accelerometer and Gyroscope Sensors (Kavitha, 2017)

Frequency-domain features were extracted.

Features	Physical Intuition	Possible Usefulness
Energy	Similar to 'energy' in the classical physics sense.	Energy of a moving activity is higher than a non-moving activity.
Entropy	The measure of consistency of the signal	Energy of moving activities are similar, but consistencies differ.
Magnitude of FT	The vibration magnitude at a particular frequency	Different modes have different vibrational modes.

Towards Clustering of Mobile and Smartwatch Accelerometer Data for Physical Activity Recognition (Dobbins, 2018)
Vibrations induced by trains: Case Study on Extension of track platform of Gare do Oriente (M. Monteiro, n.d.)
Transportation Vibration Measurement (MSU CDP, n.d.)

1. Data Cleaning
2. Fourier Transform
3. Extract features

36,592 OBSERVATIONS

734 FEATURES

367 ACCELEROMETER FEATURES

- **15** TIME-DOMAIN STATS
- **353** FREQUENCY-DOMAIN STATS

367 GYROMETER FEATURES

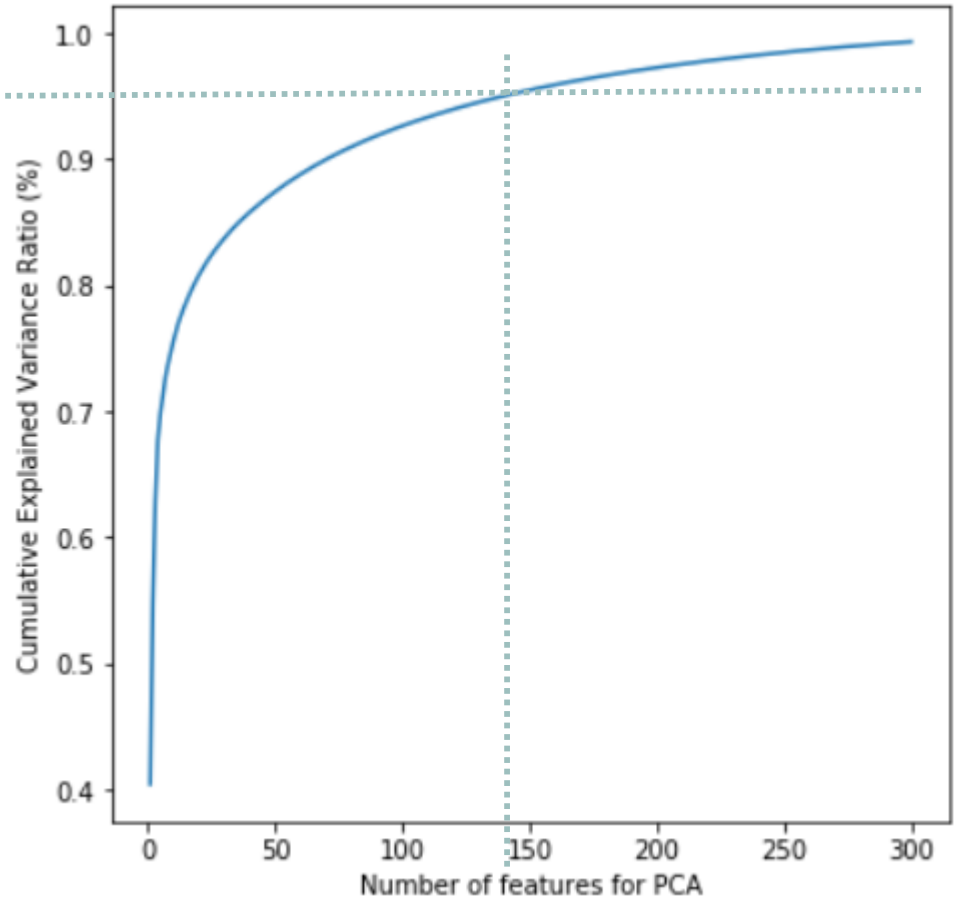
- **15** TIME-DOMAIN STATS
- **353** FREQUENCY-DOMAIN STATS

PART 3

FEATURE SELECTION

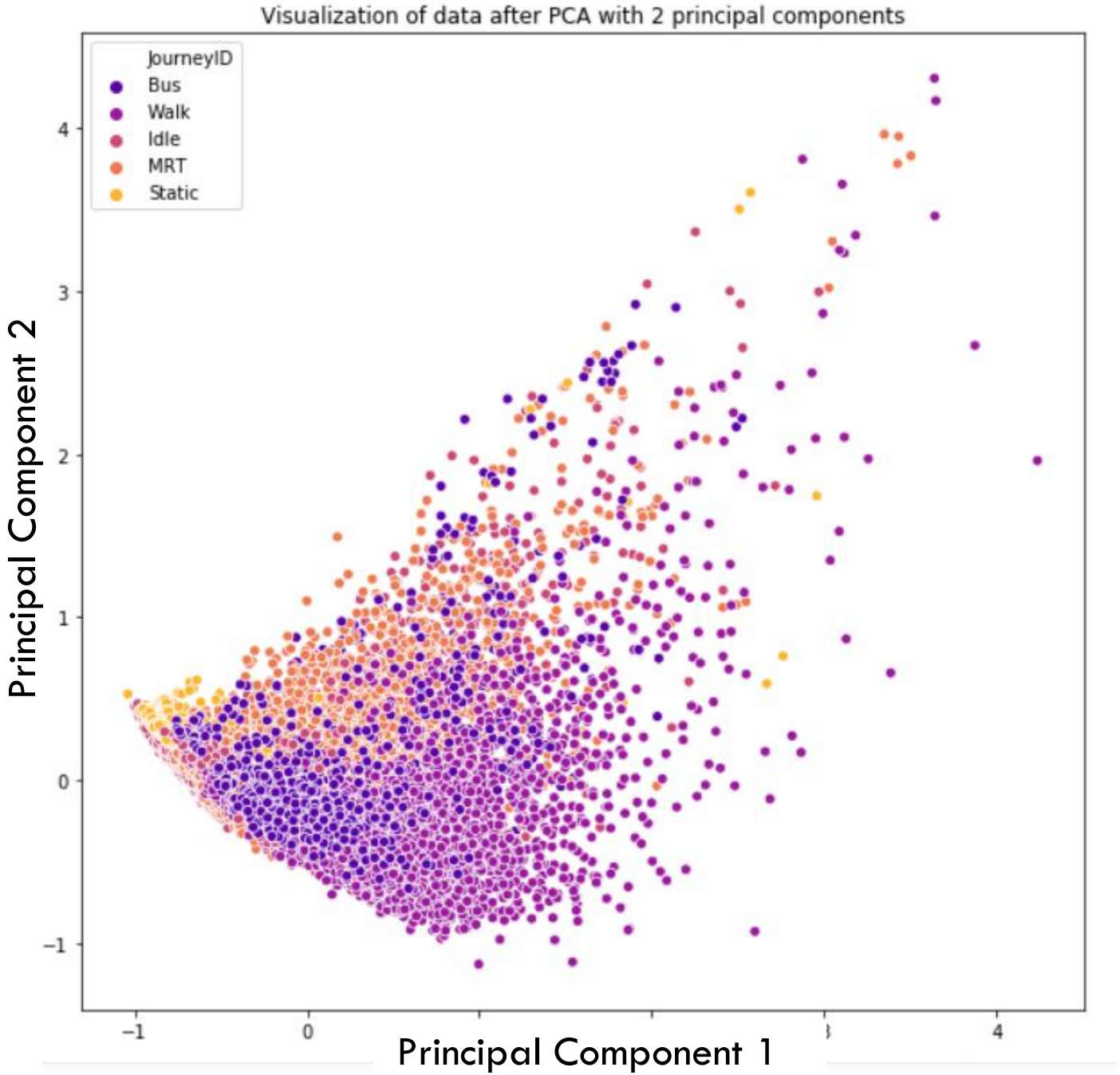
FEATURE SELECTION USING PCA

With 150 components, the total explained variance is about 95%.



DATA VIZ. USING PCA

Data is (roughly) separable!



PART 4
MODEL TRAINING

OVERVIEW

1. Split data into training and cross validation sets (90:10)

2. Train SVC Classifiers using the CV set.

3. Pick the classifier which has the highest F1-score on the CV set.

4. Predict the travel mode of CV set using the classifier.

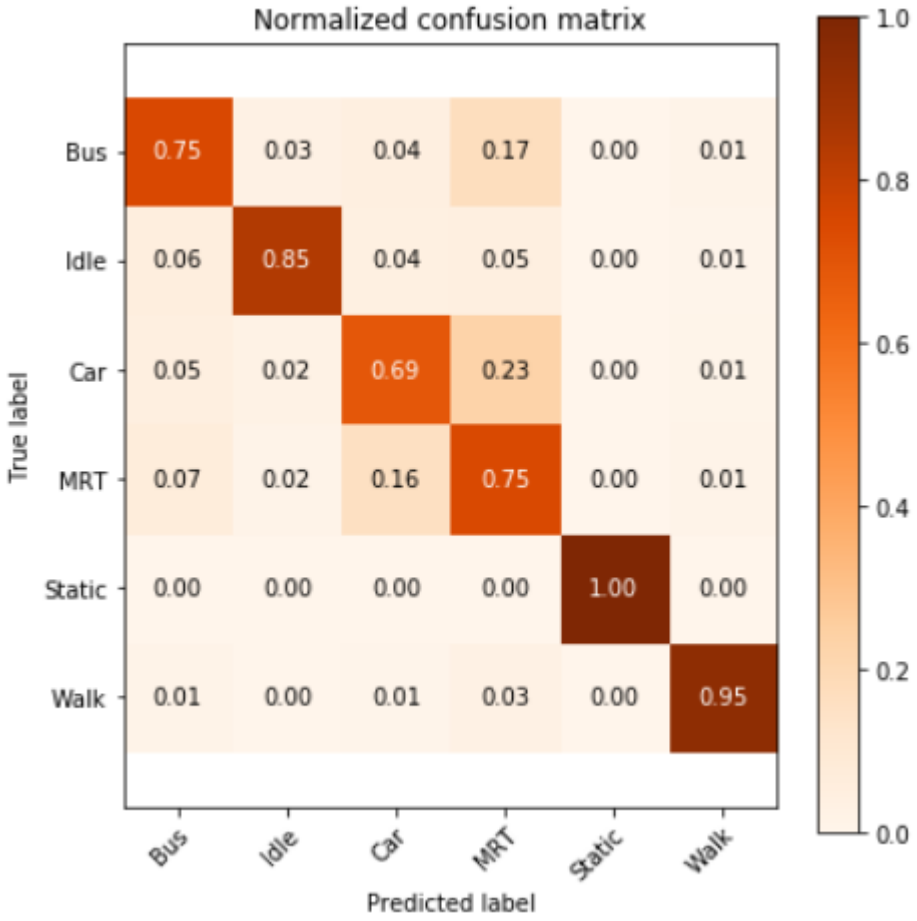
5. Evaluate the model using confusion matrix

TRAINING AND PICKING CLASSIFIER

BASE LINE	PREDICTS THE MOST FREQUENT CLASS (**MRT**) ALL THE TIME F1-MACRO = **8.7%**
MODEL SCORE	THE BEST SVC MODEL HAS F1-MACRO = **82.7%**

EVALUATING CLASSIFIER

THE GOOD	RATHER HIGH ACCURACY
THE BAD	COULD BE IMPROVED FOR CAR. MODEL COULD BE OVERFITTED.



CONCLUSION

The model presented is useful in separating different modes of transport.

No data from other phones/users are collected.

The next step would be to

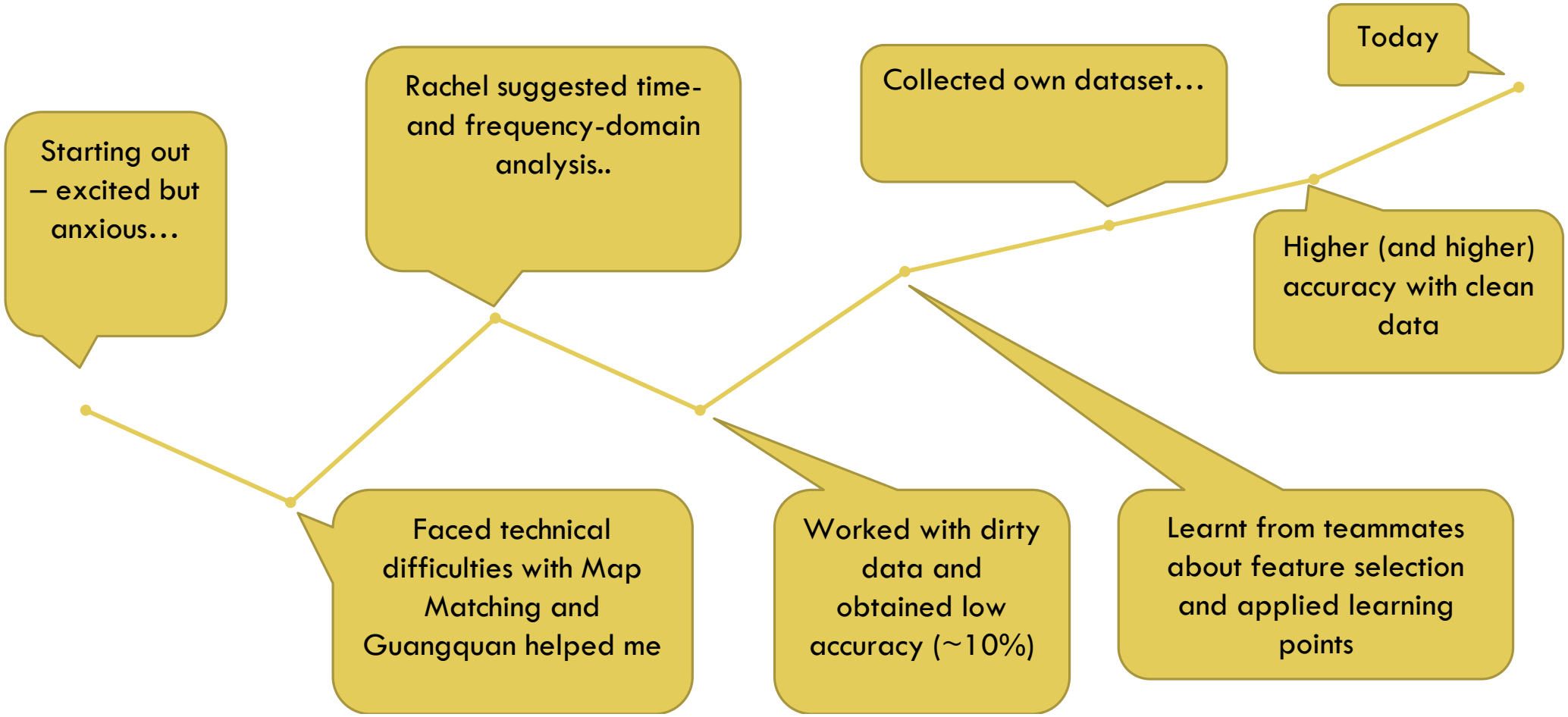
- Investigate the effect of sampling rate on model.
- Collect data for car journeys & other users.
- Select less features.

PART 5: PARTING WORDS

THE JOURNEY WITH LTA

MY JOURNEY WITH LTA

THE UPS AND THE DOWNS



SPECIAL THANKS TO…

AMAZING TEAMMATES	MS RACHEL & MR GUANGQUAN
NTU	LTA

SENSING TRANSPORT MODES WITH PHONE SENSORS

A COLLABORATIVE PROJECT BETWEEN LTA AND NTU

Presented by Voon Hao (Travis)

Guided by Ms Rachel, Mr Guangquan and Mr Jonah

28 Feb 2020

Land Transport Authority