Hubway

*Data-driven market segmentation to draw
insights for a coherent user expansion strategy*

2nd March 2022

Adelyn | Xin Jie | Shannon | Jun Li | Jiadong

Executive Summary

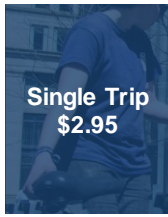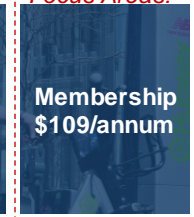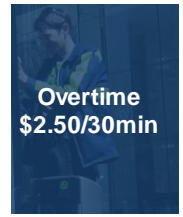|	
Introduction	• Users of Hubway pay a membership fee which includes unlimited trips up to 30 minutes in duration at no additional cost, with additional "overtime" fees charged for trips exceeding 30 minutes
Objective	• Management has asked to conduct **market segmentation** to learn more about the customers' profiles and characteristics, and **ways to expand its user base**
Marketing Analytics	• **Who to target?** – Creating customer personas based on k-means clustering • **Where and when to reach out?** – Data based on k-means clustering and geospatial data retrieved from Hubway • **How to attract and engage?** – Marketing strategies based on main target audiences
Model Approach	• **Exploratory data analysis** coupled with objective assessment to identify potential relationships between variables • Choosing of optimal k using **Elbow method** • Uses all variables in k-means clustering
Model Evaluation	• Cluster plot uses **Principal Component Analysis** to understand how the clusters relate to each other and the different variables • Cluster plot shows that there are **differences between most** clusters apart from two clusters with some similarities. • Results of K-Means clustering allows us to identify 3 different customer profiles
Customer Persona	• **Office workers** (Groups 1 – Morning, 3 – Afternoon / Lunch Break, 4 – Evening) • **Night Cyclists** (Group 2) • **Weekend Leisure Cyclists** (Group 3)
Limitation	• Data provided only contains four independent variables that potentially affect a trip's duration • Analysis of each trip's start and end station and member type would be useful for deeper analysis – additional data was retrieved from Hubway's website
Recommendation	• **Prioritise** marketing efforts on **office workers & weekend cyclists** as they are unlikely to own a bike, frequent riders and favours convenience • Marketing efforts reaches out to their **common touchpoints** and addresses their need for **convenience and long-term value** • F2F marketing efforts can be held at Downtown Boston where majority of Hubway's target audience and rides are

Problem Overview and Objectives

Our strategic marketing plan for growth-focused Hubway…

1 Business Overview

Founded in Boston in 2011, Hubway is a bike sharing company. Users can borrow a bike from 1 of 140 stations and return to a different station. It is looking to **increase the number of memberships on its platform.**

Focus Areas:

Single Trip $2.95	Membership $109/annum	Overtime $2.50/30min

Membership-focus drives flywheel of growth

- Improve marketing and service offerings
- Increase memberships
- **Sustainable Growth**
- Increase Recurring Revenues

2 Our strategic growth plan

Growth-hacking by focusing, localising and executing

Who to target?
Focus

Where and when to reach out?
Localise

How to attract and engage?
Execute

… is data-driven and insightful

3 Why k-means clustering?

| Advantage	Why is it important?
Simple Implementation	K-means can **scale to large datasets (as per our dataset)** and be easily executed
Generalises to clusters of different sizes	In segmenting our customer base, k-means generalises well to appropriate clusters, allowing us to **conduct profiling and identify key characteristics within the cluster**. Filtering different cluster sizes **help in prioritisation**.
Easily adapt to new examples	**Allows model iterations** as Hubway grows and adjusts to their next phases of growth

4 Our systematic approach

Our data-driven approach will deliver the most benefits to Hubway

Play to Hubway's strengths	Understand Hubway's value-add	Sell a coherent Hubway story
Unlock source of organic growth previously identified	*Ride on how Hubway drove value for customers*	*Return high marketing ROI by converting target audience*
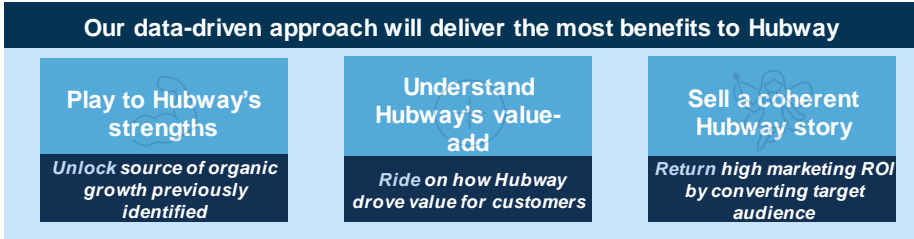
Steps taken:

- K-means clustering
- Problem analysis

- Cluster analysis
- Profiling
- Identify key growth levers

- Geographical Analysis
- Differentiated marketing strategy

Exploratory Data Analysis (Univariate)

Trip Duration (Outliers Removed)

Skewness of Duration

Before	After
130.07	1.204
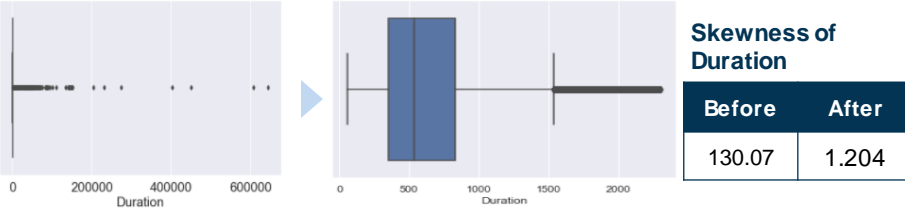
Removing Outliers using logic and IQR[1] technique
- Lower bound cut-off: 1 min – account for false starts or redocking of bikes
- Upper bound cut-off: 3 * IQR – account for trips that are too long (lost bikes)
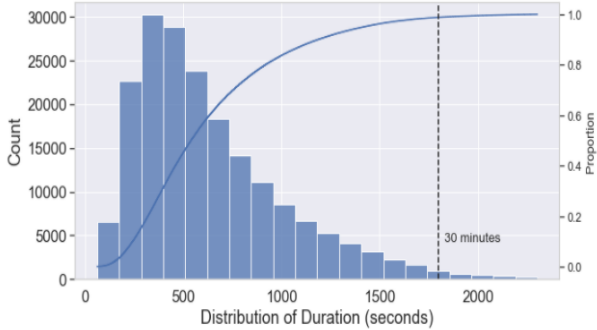

- Very few trips (< 5%) incur "overtime" charges as most people ride below 30 minutes

- Trips mostly concentrated around the 400 second mark (5 - 6 mins)

- > 90% of trips are within 30 minutes

Hypothesis
- Bikes are rented for commuting between short distances rather than long, leisurely rides

1. IQR: (Inter-quartile range) spread difference between the 75th and 25th percentiles of the data

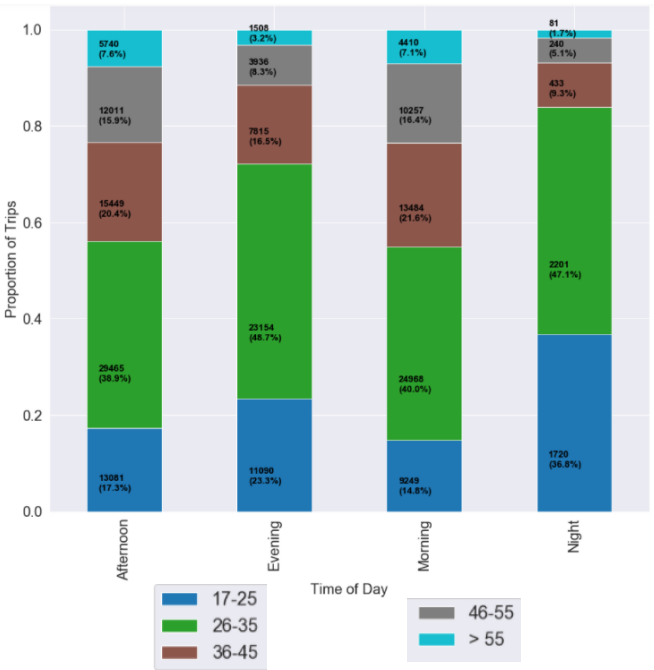## … and understanding how time of day and age matters
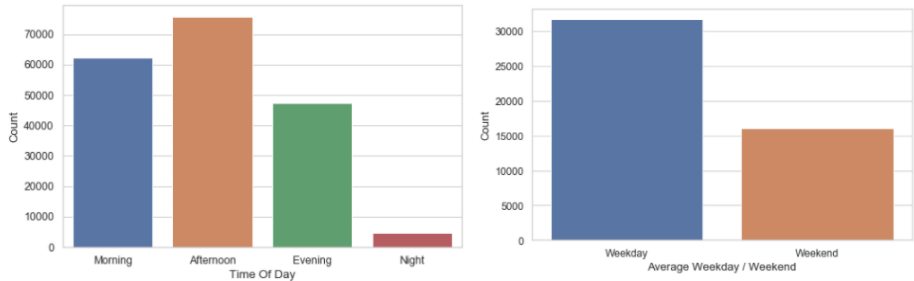
- Trips are most popular in the afternoon, then the morning and evening
- On average, there are twice as many trips on weekdays than weekends
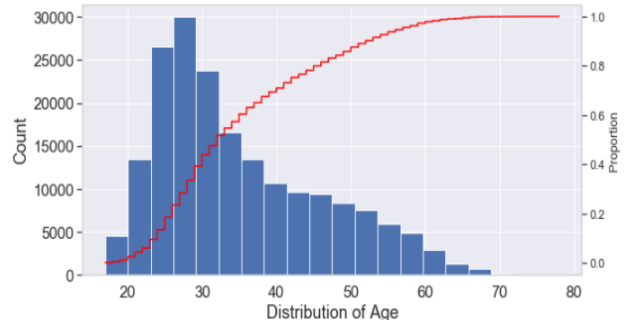
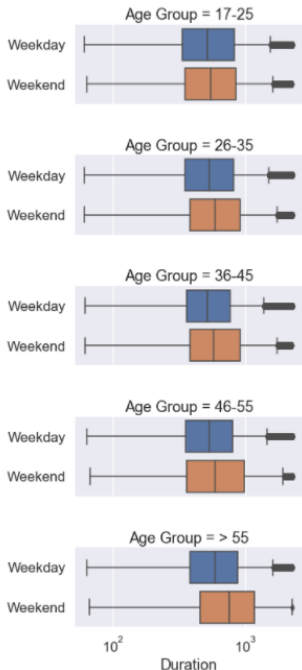- Most riders are younger individuals aged between their 20s and 40s

# Exploratory Data Analysis (Multivariate)

Age Group vs Time of Day

- The proportion of teenagers between the ages of 17-25 who rent bikes increase at night (36%), as compared to the day (14-23%)
- Older individuals above 55 make up a very **negligible proportion of trips all day and night** (< 8%).
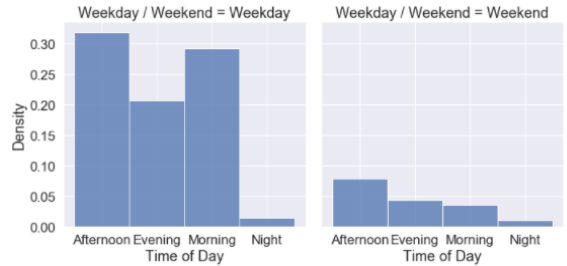
Trip Duration

- Trip durations largely stay the same between weekdays and weekends
- Older individuals above 55 have a **higher median trip duration on weekends than weekdays**

Trip Density by Age Group and Time Period
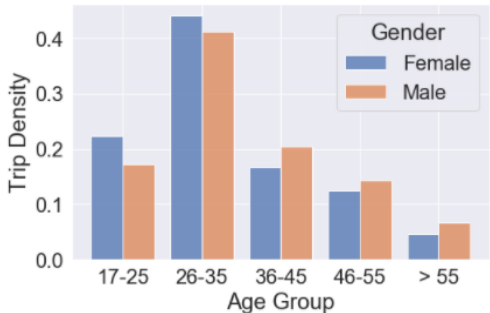
Ages 17 – 35: More females ride bikes than males
Ages 36 and above: More males ride bikes than females

- On **weekdays, trips peak on afternoons and mornings**, followed by evenings and night.
- On weekends however, trips peak on afternoons, followed by evenings and mornings.

Exploratory Data Analysis (Multivariate) and Strategies

Evaluating effects of gender on users

Trips breakdown by Gender

MAMIL (Middle Aged Men in Lycra) Social Phenomena
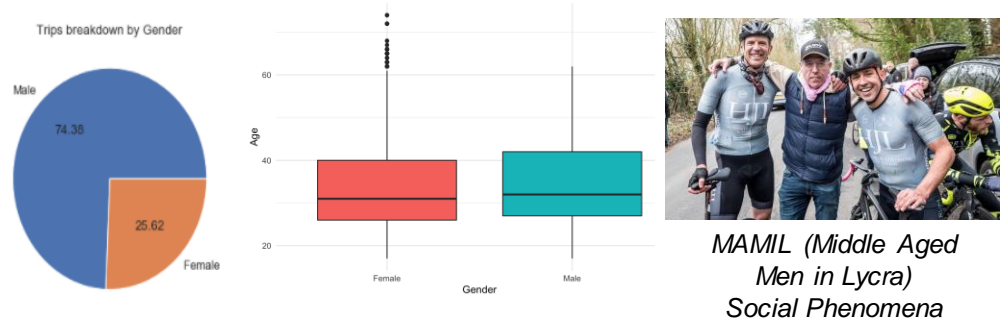
Hypothesis
- Also validated by literature[1], **middle-aged men are more likely to utilise bike sharing apps** and hence, is an important customer segment
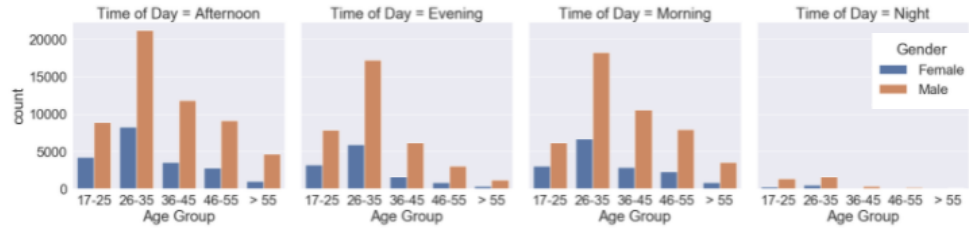
Trip trends between males and females are largely similar for different times of the day and age groups.

Hypothesized strategies based on EDA

*Consolidated Insights and Potential Targeting Strategies*

Time of Day	• **Prioritize morning and afternoon timings on weekdays for maximizing availability of bikes**, so that users can easily locate bikes • Trips at night can be made cheaper • Bike availability on weekend afternoons can be increased
Age Group, Gender	• More younger females ride bikes compared to males, but this trend reverses when age increases. • **Differential marketing appeal by introducing more "female" themed bikes** to appeal to the younger female population i.e. Hello Kitty Themed Bikes • Bikes should be designed to appeal to the segment in the 26-35 age group • Older individuals > 55 are not bulk of our users, hence might not need to be targeted
Trip Duration	• Majority of users make short trips (<30 mins), offer more attractive packages for shorter trips • Consider shortening unlimited package to 20 mins to increase availability of bikes (more users can use)

Source(s): 1. Journal of Sport and Social Issues (Why We Ride: Road Cyclists, Meaning and Lifestyles)

Problem Analysis	**Exploratory Data Analysis**	Cluster Analysis	Recommendations	Conclusions

Model Optimization: Elbow Method

Finding the optimal number of clusters

Why is it important?
Too many clusters can lead to artificial boundaries within real data clusters

Plot of # Clusters vs. tot.withinss

Elbow – K = 5
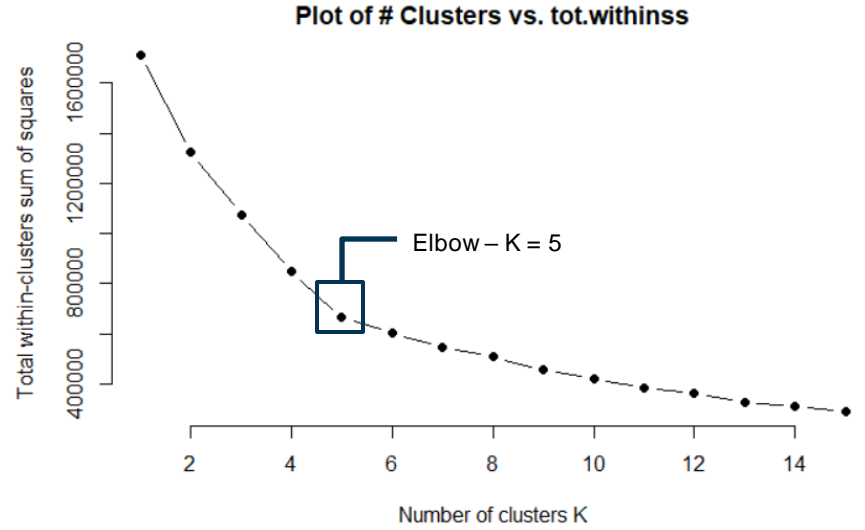
The **elbow method** plots the within-cluster sum of square (WCSS) between different cluster values. The location of the bend in the plot indicates the optimal number of clusters for the model.

Evaluating the clusters

Cluster plot uses principal component analysis (PCA) to visualize the clusters in two dimensions
Top 2 Principal Components explain 40.9% of point variability
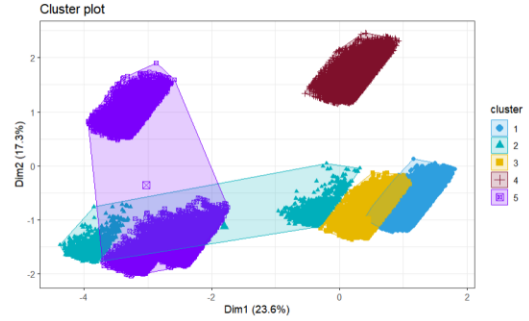
- There is **minimal to no overlap** for most clusters which indicate that there are **significant differences** in the groups identified.
- On the other hand, **Clusters 2 and 5 shares some similarities.**

More in-depth Visualization of All Principal Components
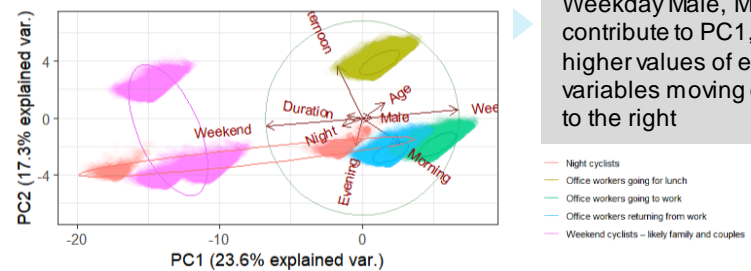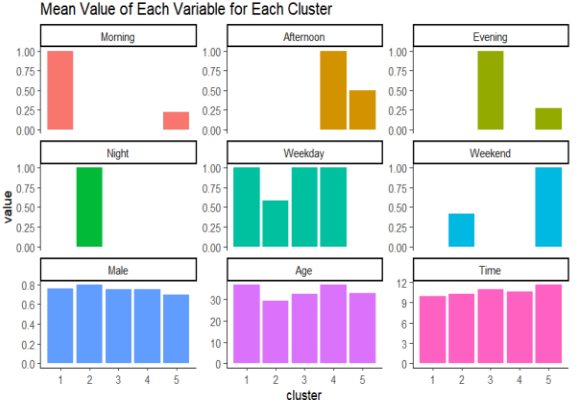

Here, we can understand **how each cluster relates to each axis**. E.g., Age, Weekday Male, Morning contribute to PC1, with higher values of each variables moving clusters to the right

Model: Identification of Clusters

K-means clustering returns us 5 differentiated customer profiles

Group.1 <int>	Duration <dbl>	Morning <dbl>	Afternoon <dbl>	Evening <dbl>	Night <dbl>	Weekday <dbl>	Weekend <dbl>	Male <dbl>	Age <dbl>
1	595.1496	1.00000	0.0000000	0.0000000	0	1.000000	0.000000	0.7559237	36.94624
2	617.4875	0.00000	0.0000000	0.0000000	1	0.582246	0.417754	0.7989305	29.53390
3	658.1300	0.00000	0.0000000	1.0000000	0	1.000000	0.000000	0.7512665	32.70914
4	632.4293	0.00000	1.0000000	0.0000000	0	1.000000	0.000000	0.7474967	37.08854
5	698.2020	0.22704	0.4996342	0.2733258	0	0.000000	1.000000	0.6953515	32.91607

Mean Value of Each Variable for Each Cluster

Distribution of User Duration by Cluster

Distribution of User Age by Cluster

Observations
- Gender, age and duration of bike borrowing is similar across different group profiles
- **K-means profiling is done based on time and weekday / weekend**
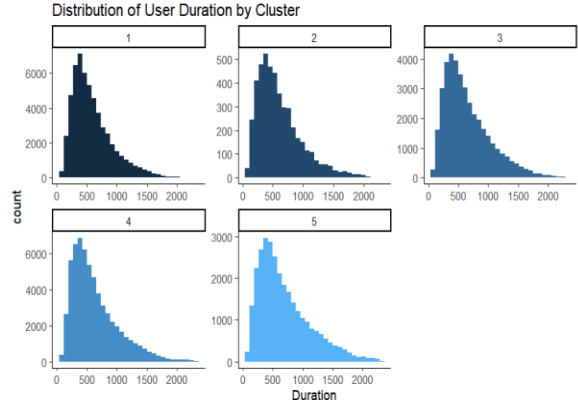
Observations
- The distribution of duration of bike borrowing is similar across different group profiles
- **No specific group that has a "high-value" segment that should be prioritised**
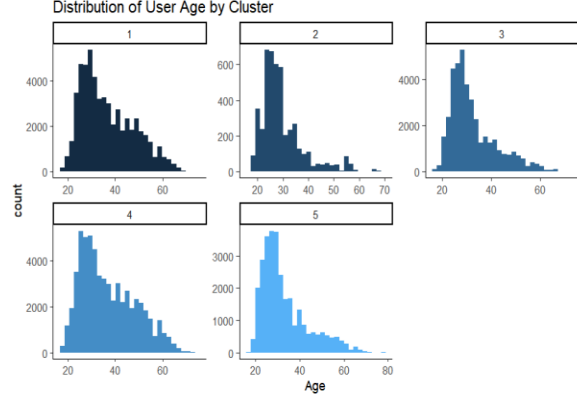
Observations
- The distribution of user's age is similar across different group profiles
- Other than Group 4, which are mainly 20s-30s, other groups have a **concentrated user profile of 20 - 50**

Notes: 1. Initial hypothesis of target persona, will be validated in subsequent slide; 2. Age will be further explored in subsequent slide

Cluster Analysis – differentiating clusters and identifying customer profiles

K-means clustering returns us 5 differentiated clusters that can be summarized into 3 main customer profiles

	Group 1	Group 2	Group 3	Group 4	Group 5
Target Persona	Office workers going to work	Night cyclists	Office workers returning from work	Office workers going for lunch	Weekend cyclists – likely family and couples
Distinctive Attributes	Morning = 1.0 Weekday = 1.0	Night = 1.0 Weekday = 0.6 Weekend = 0.4	Evening = 1.0 Weekday = 1.0	Afternoon = 1.0 Weekday = 1.0	Weekend = 1.0 Morning = 0.23 Aftn. = 0.50 Evening = 0.28
Gender	Male = 0.7559 ~equal gender ratio	Male = 0.7989 Slightly more males	Male = 0.7513 ~equal gender ratio	Male = 0.7475 ~equal gender ratio	Male = 0.6954 Slightly more females
Mean Age	36.9	29.5	32.7	37.1	32.9
No. of users	55,540 (28.6%)	4,675 (2.41%)	39,283 (20.2%)	60,720 (31.3%)	30,074 (15.8%)

# of Users in Each Cluster
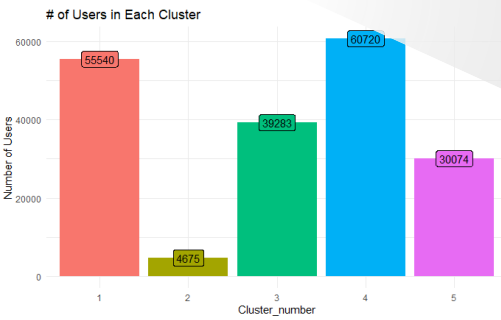
Customer Profiles:

Office Workers

Night Cyclists

Weekend Cyclists

Legend:

Main Target Segments

Night cyclists have a low user base and are unlikely to adopt Hubway's services[1]

Source(s): Team Analysis; Notes: 1. Explained in next slide

Marketing Strategy: Understanding that our target groups have different uses for Hubway, our marketing strategy will be differentiated across target groups

Legend: [Main Target Segments]

Our target groups can be split into three distinct personas, **each with their own uses for Hubway's bikes,** meaning we need a **differentiated marketing and monetisation strategy for each target group**

	"Health-conscious office workers" High-income working adults	**"Leisure weekend riders"** Families and couples	**"Avid night cyclists"** Enthusiastic young adults
Targets & Personas			
Market Penetration	TAM — *TAM : 418,890* / *TAM Captured: 37%* / *User No.: 155,543*	TAM — *TAM : 418,890* / *TAM Captured: 7%* / *User No.: 30,074*	TAM — *TAM : 418,890* / *TAM Captured: 1.1%* / *User No.: 4,675*
Cycling habits	↑ **Frequent and habitual**, most likely as connecting rides to public transport systems	— **Infrequent and occasional**, most likely during special family events or outings	↓ **Highly frequent**, most likely as part of a cycling group at their own availabilities
Consumers' Preferences	Bike accessibility ▮▮▮▮▮ / Bike quality ▮▮	Bike accessibility ▮▮▮▮ / Bike quality ▮▮▮	Bike accessibility ▮ / Bike quality ▮▮▮▮▮
Bike ownership	↓ **Less likely** to own, daily use of bikes limited to short time intervals, use other transportation	— **May** own if depending on family use of bikes, but unlikely high ownership across families	↑ **Very likely** to own, prefers sense of ownership of quality bikes for regular night cycling
Likelihood of adoption	**Highest likelihood**, attracted into Hubway's key value proposition of convenience	**Probable**, but infrequent customer base – membership takes some convincing	**Unlikely**, but potential to attract the rare, few cyclists based on convenience

Priority level

Problem Analysis	Exploratory Data Analysis	Cluster Analysis	**Recommendations**	Conclusions

Marketing Strategy: For Hubway to succeed, gaining traction through early adoption is key – this means rapidly pushing target customers through the marketing funnel to eventual conversion and beyond

Marketing Funnel
Differentiated approaches for each target group to **rapidly secure early adopters**

Health-conscious office workers

Leisure weekend riders

From customers to evangelists

Awareness
Becoming known in the market will be a challenging first step…

Engagement
…and engagement will keep them interested…

Conversion
…before we make the final push for conversion

Targeting office workers' common touchpoints
Company promotions (B2B2C), Social media

Targeting leisure riders common touchpoints
Family / Couple package offers, Community events, Household Visits

Targeting office workers' need to see convenience
Offer free-trials and ensure bike availability at popular hotspots[1]

Targeting the office workers' need to see long-term value
Increase bike rental stations, family events to promote other uses

Hubway can co-organise with communities grassroot cycling events to change the frame of mind as a **"commuting bike" to a "lifestyle subscription"**.

Hubway can offer **discounts to current users to encourage** viral signups amongst families and peers

Notes: Popular hotspots determined in the subsequent slide

| Problem Analysis | Exploratory Data Analysis | Cluster Analysis | Recommendations | Conclusions |

Marketing Strategy: Hubway can concentrate its in-person marketing strategy at Downtown Boston, where the main target audience with high conversion chance commutes to

Heatmap depicts majority ridership in Boston's CBD…
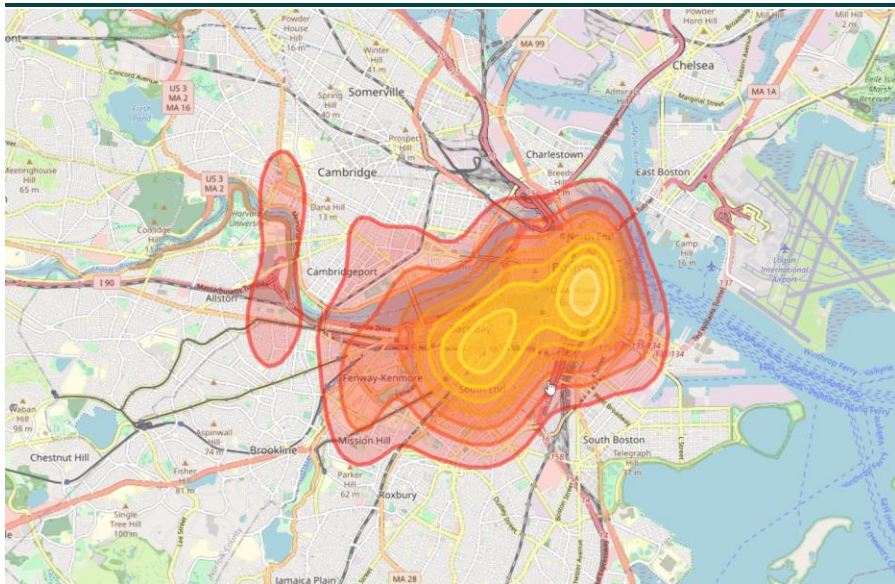
… thus, focus on top 10 stations/routes drives impact
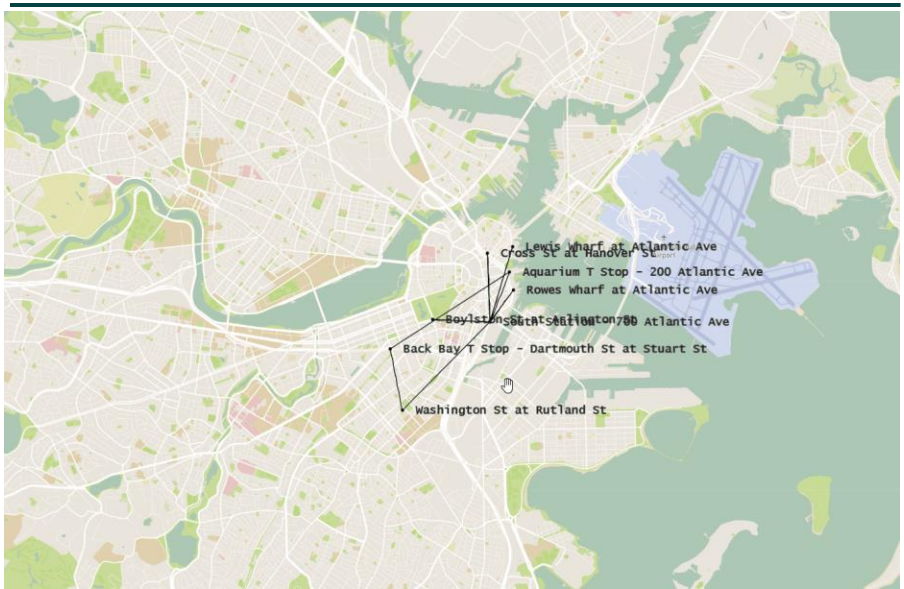
- Heatmap of 2012 rides demonstrate that majority of Hubway's rides are localised around **Downtown Boston**, which is **considered the central business district** within Boston
- Validates our thesis that **ridership mostly comes from office workers,** and should **double down marketing efforts there**

- In-person marketing (e.g. flyers, promotions, free-trial campaigns) can be carried out in the morning, afternoon and evening at the top 10 bike stations/routes
- **High ridership at these areas serves as a "product demo" and reinforces positive brand image amongst on-the-fence office workers leading to user expansion**

Source: Bluebikes Hubway Trips (2012) csv files,

| Problem Analysis | Exploratory Data Analysis | Cluster Analysis | Recommendations | Conclusions |

Conclusions

Risk and Mitigation, Further Recommendations

1 *Critical Risk Analysis*

	Risk	Mitigation
α	**Collaboration** Private companies might not be keen to collaborate	Focus on companies with high Hubway users to build credibility and trust
β	**Financial** Larger than expected marketing costs with initiatives	Budget mgmt., extend duration to rollout marketing initiatives
3	**Referrals** Incentives proposed are not attractive, do not induce virality	Seek and iterate on feedback, adjust Hubway offerings continuously

2 *Further areas of opportunities*

Growth Opps.	Analysis
Monetising short rides	Average ride duration of office workers (Group 1, 3, 4) is ~9 – 11 mins, this means that one-way rides (e.g. home to office) is preponderant[1] and **overtime accounts for little revenue as a whole[2]**. Hubway can **charge fees for bike returns at a different station than the pick-up point** at a later growth stage to **further monetise their current base of customers**.
Return Policy	We noticed that a small minority[3] of riders fail to return their bikes in 1 day. **Hefty overtime charges and fines should be applied, especially if bikes are lost or stolen**.

Notes: 1. 53% of dataset, 2. 90% of dataset, 3. 0.01% of dataset

Model Limitations and Lifecycle Management

3 *Disadvantages of k-means clustering*

Disadvantages	Why?
Choosing k manually	Even though the optimal k is selected by the elbow method or silhouette analysis, **more clusters could mean more granular segmentation and better customer profiling**. In our case, our clusters were formed based on time or weekday/weekend with limited differences in gender or age. In a larger dataset, information could be lost with a smaller k.
Clustering outliers	Filtering was done before k-means clustering. However, the presence of outliers could lead to **misinterpretation of a target segment** when there is in fact, none.

4 *Model lifecycle and management*

Objective	• Capture new data for continuous learning • Retrain models so they continually adapt to the dynamically changing customer segments of the company
Benefits (Context-ualised)	• Prioritise customer segment and marketing focus • Track effectiveness of marketing strategy on target demographic and enables A/B testing
Future consider-ations	• Adopt the use of alternative data – competitor data, credit card data (when signing up as a member), geospatial data to further segment the customers by income and location