Customer stream improvement proposal

Action commitment analysis with Clustering and Applied Scientific approach for Capital Bikeshare's business

Business Analytics 20595

Bocconi University, MS in Data Science and Business Analytics, 2021 – 2022 a.y.

Group 3
Chiavarino Federico, Giancola Simone, Liscai Dario, Marchetti Simone

capital bikeshare

General info

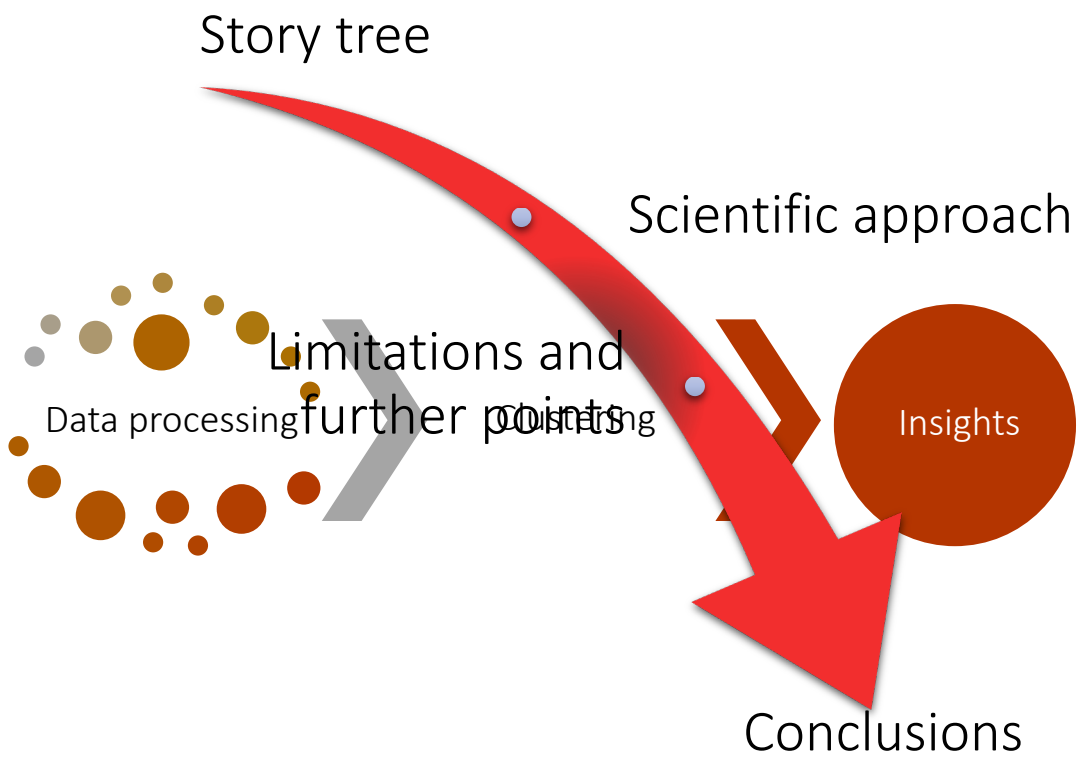
Capital Bikeshare is metro DC's bike share system

Born in September 2010

Operated by Motivate

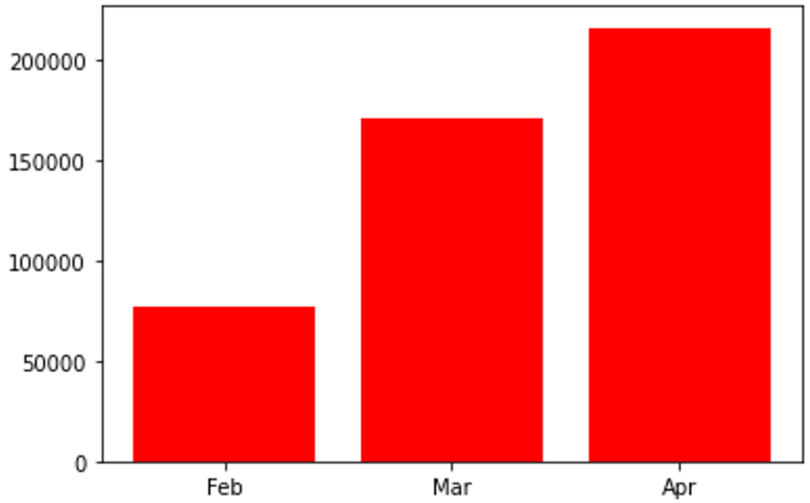
Aims

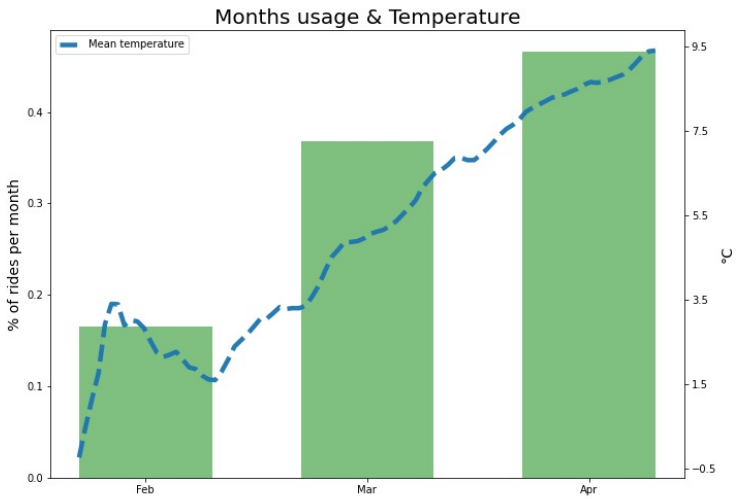
- Exploit weather and rides information to design a procedure that benefits the company
- Follow a rational line of reasoning to return a robust result
- Bring evidence to support claims made along the process
- Draw reasonable conclusions and assess limitations

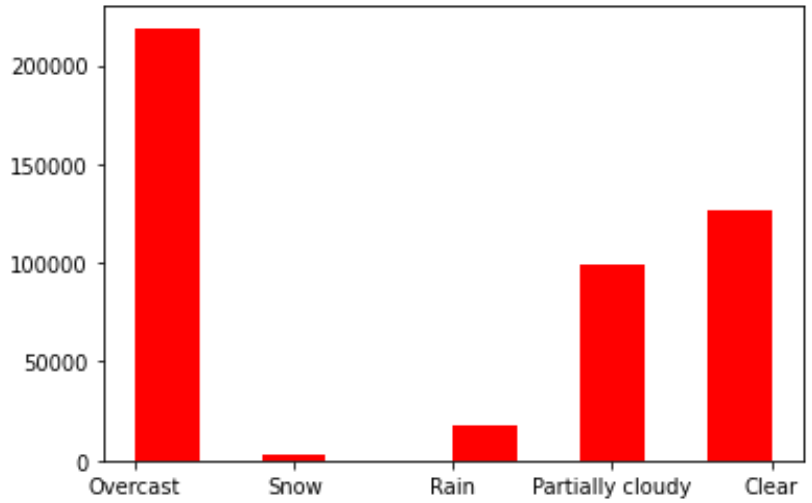


Descriptive Statistics

Customers are heavily influenced by weather conditions and temperature when deciding to use Capital Bikeshare's bicycles







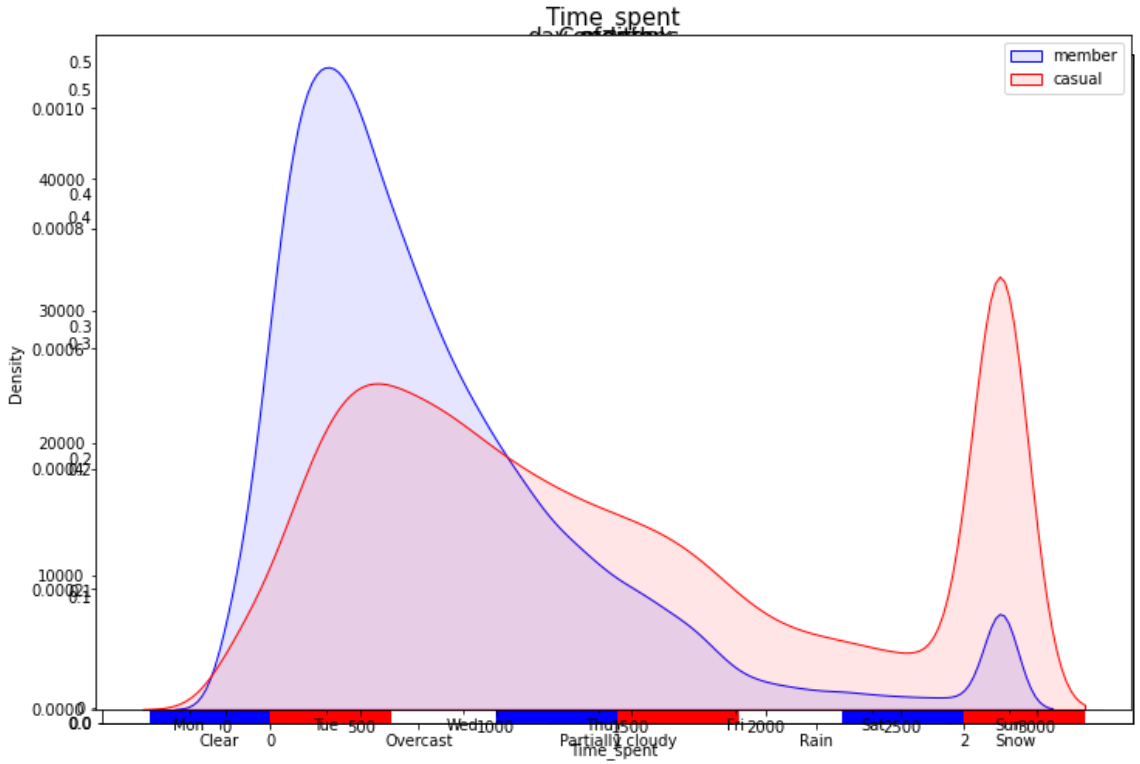
Descriptive Statistics
members vs casuals

Analysis

Members: annual membership
Casuals: single ride / day pass

58% members, 42% casuals

Comparison members vs casual customers identifies stronger differences across data. This happens in all the features we considered



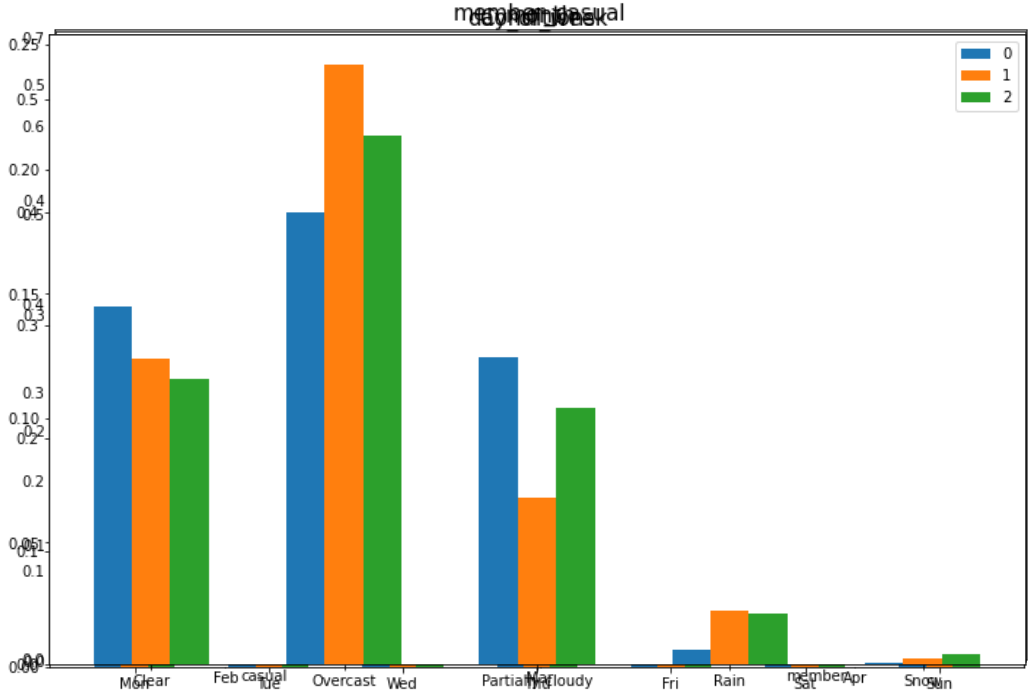
Clustering

Analysis

Cluster analysis to identify distinct groups of rides with similar characteristics

K-Means: 3 centroids

- Deeper analysis of the rides: discover patterns
- *Group 0* seems to use the service mainly with favourable conditions. In fact, the usage on
- The usage is higher than the other two groups, attaining around 65% of the rides happening in April only
- There are just slight differences between



Story Tree

Tools that enable us to derive promotions in a logical way of thinking

Identify your story

What am I observing? Different types of rides

Why is that? Different customers' needs

What can I do? Exploit these differences to craft new promotions

Identify most important elements & reasons

Types of rides

- casual
  - Flexibility
  - Price
- members
  - Specific constant needs

Current offers

- Impact on being member or not

Differences

1. Time spent
2. Conditions
3. Days of week
4. Months
   1. Leisure vs commuting
   2. Have fun vs urgent needs
   3. Free time vs routine
   4. Leisure vs constant needs

Other companies' offers

- New insights + understand differences

What could we offer?

- Flexible promotions in order to suit customer needs
- Promotions + other facilities

Scientific Approach

Step 1: Idea

Different usage of service depending on several factors

Promotions in terms of flexibility of usage and possible partnerships

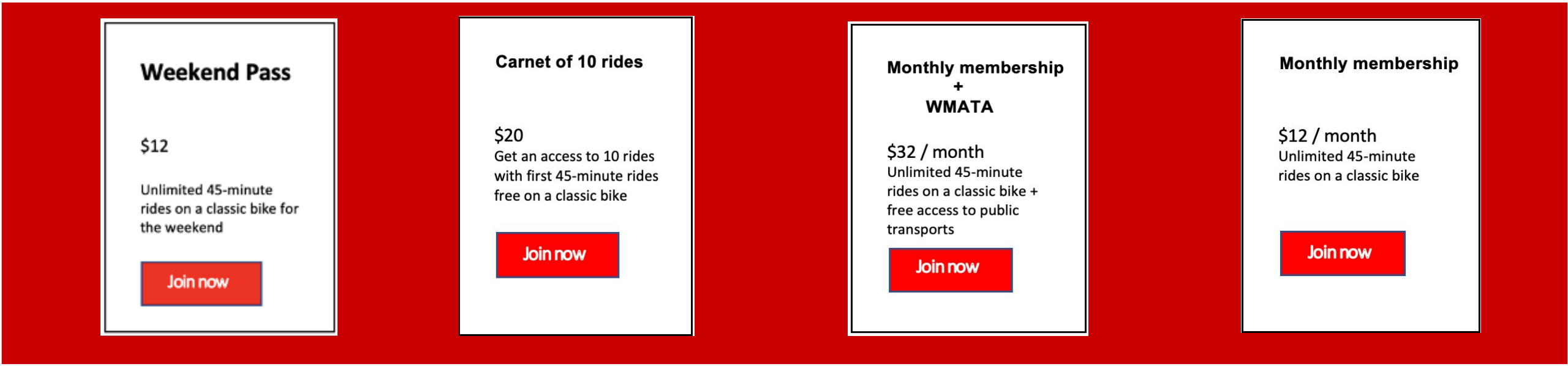
Step 2: Problems

Casual customers

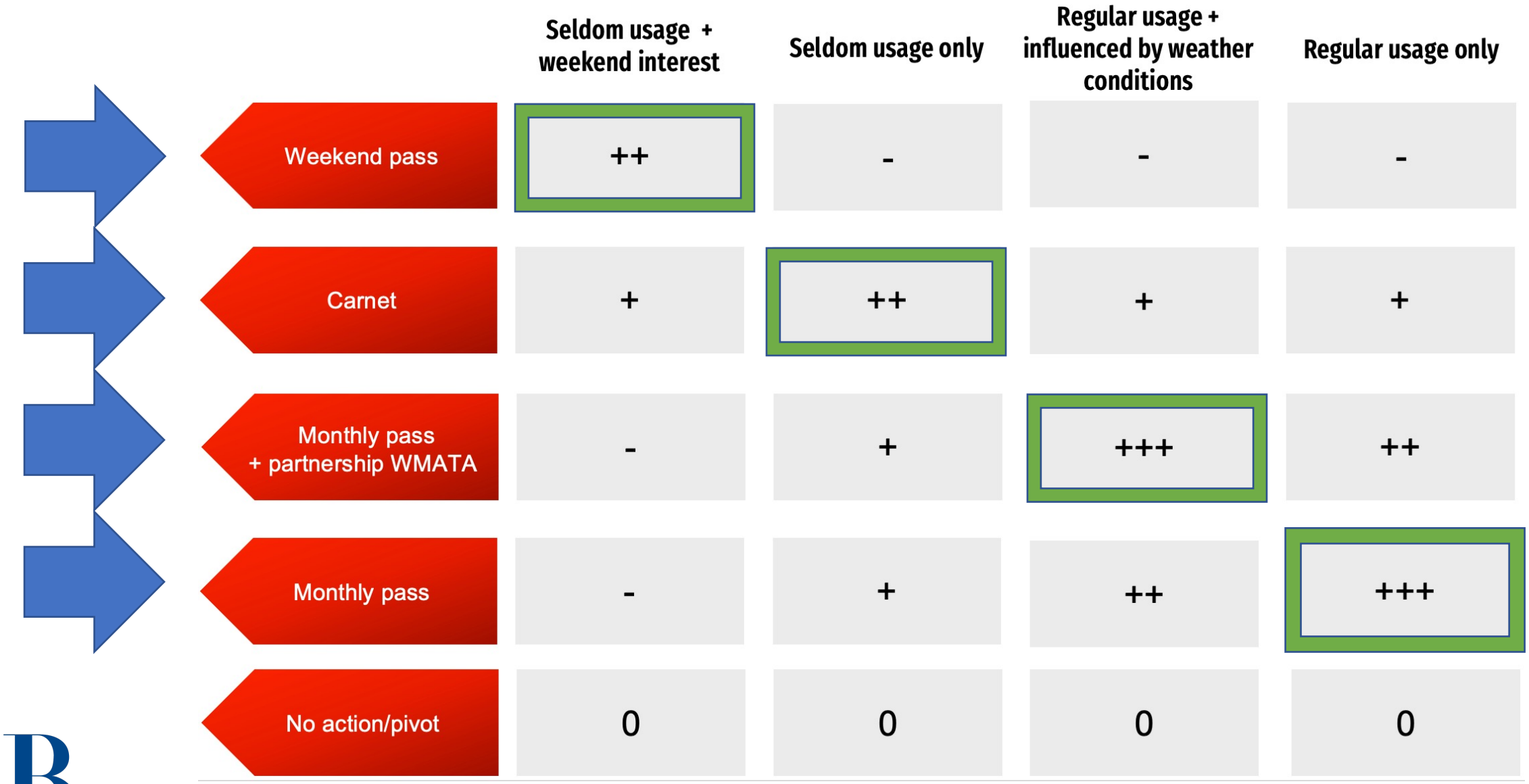
Revenue drops during coldest months and days with bad weather conditions

Step 3: Scenarios

Step 4: Actions

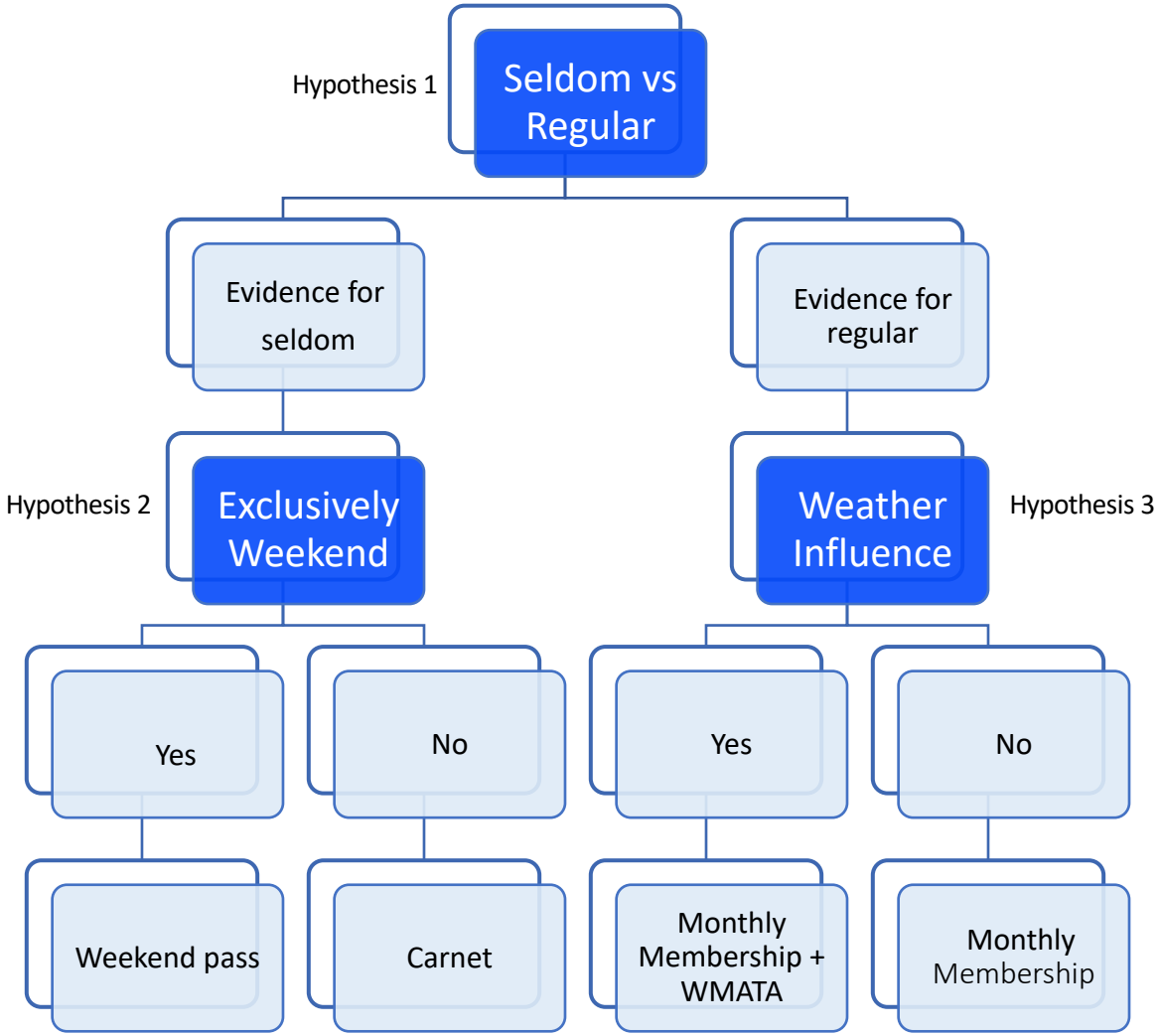


Step 5: Scenario Action Map



	Seldom usage + weekend interest	Seldom usage only	Regular usage + influenced by weather conditions	Regular usage only
Weekend pass	++	-	-	-
Carnet	+	++	+	+
Monthly pass + partnership WMATA	-	+	+++	++
Monthly pass	-	+	++	+++
No action/pivot	0	0	0	0

Step 6: Testable Hypotheses



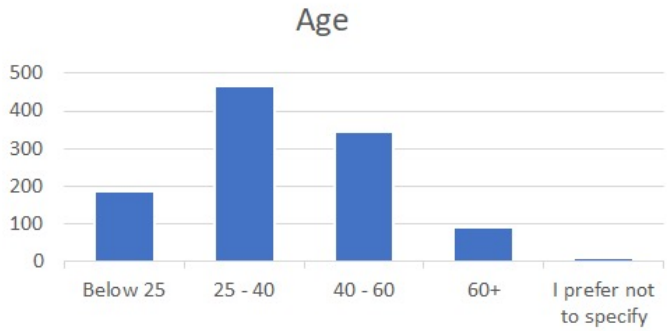
H0: The majority of customers uses the service in a regular way

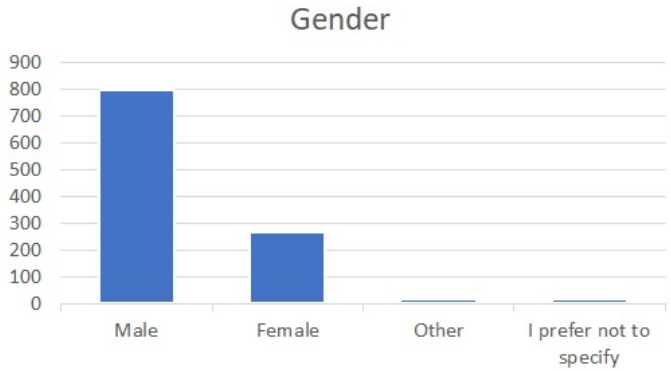
H0: the majority of sporadic users does not have an exclusive interest in weekends

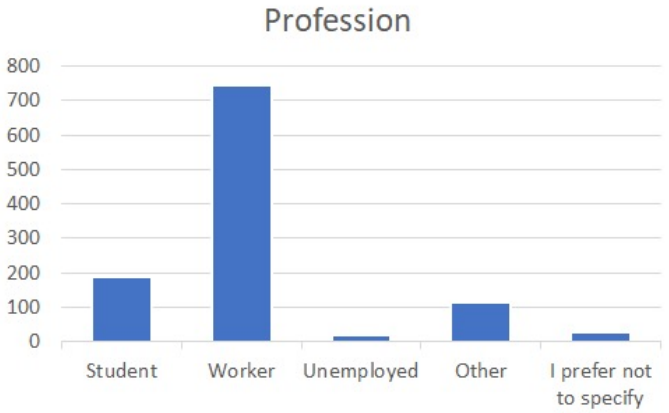
H0: people are not influenced by the weather

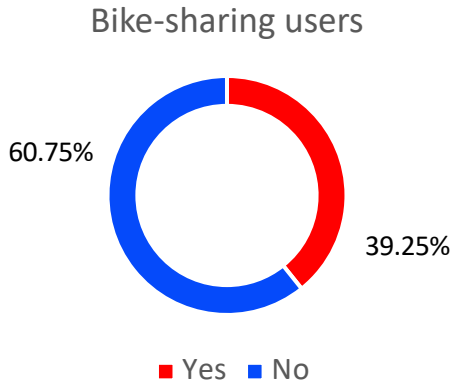
Survey Statistics

1093 respondents









Survey Procedure

Ideal Target: pool of customers and potential ones

Our target

Cycling's Subreddits

↓

Mimic Capital Bikeshare's customers

General individuals

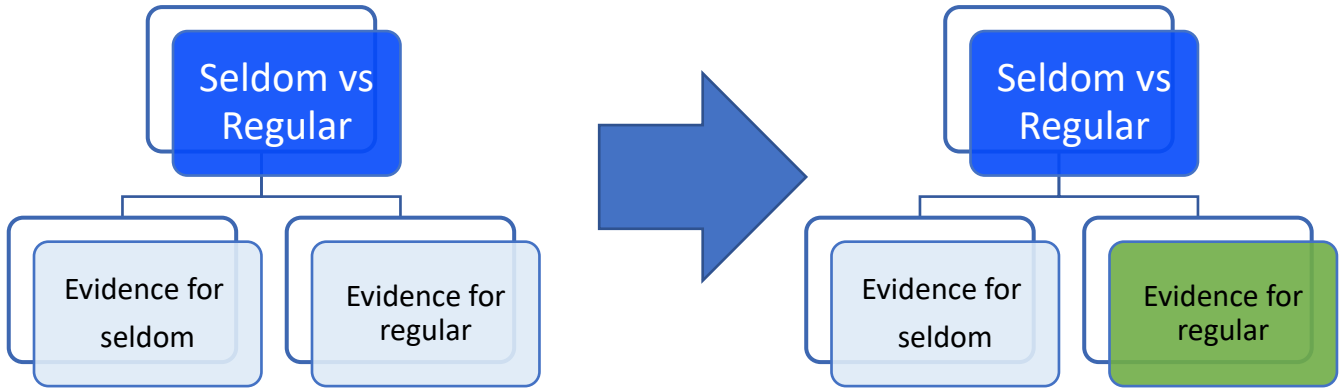
↓

Potential customers

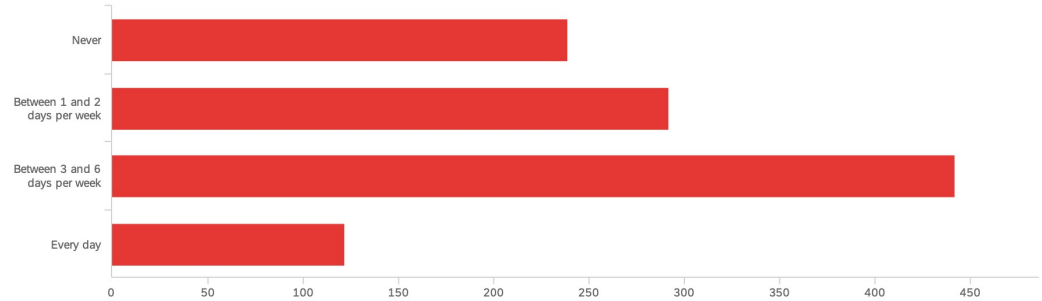
Step 7: Testing the results

Survey's questions

H0: The majority of customers uses the service in a regular way



Q2 - On average, during any week, how often do you use a bicycle?



Stata's test

One-sample t test

Variable	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
regular	1,093	.5141812	.0151246	.5000276	.4845046	.5438577

mean = mean(regular) t = 0.9376
H0: mean = .5 Degrees of freedom = 1092

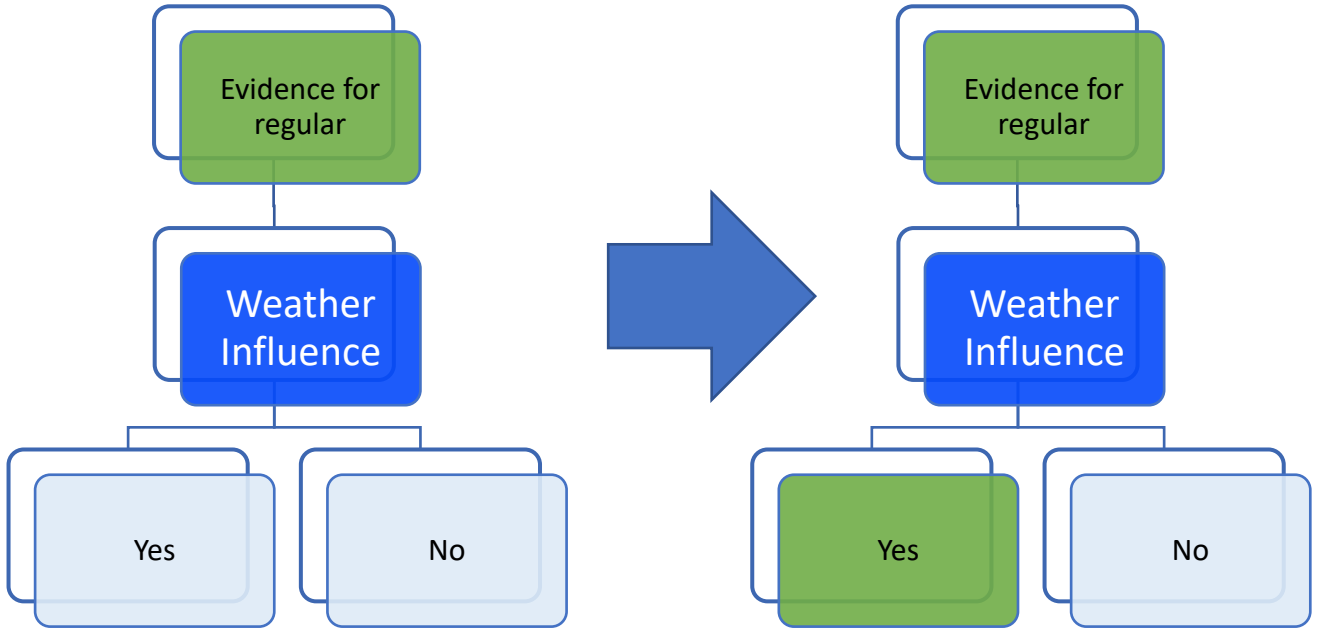
Ha: mean < .5
Pr(T < t) = 0.8257

Ha: mean != .5
Pr(|T| > |t|) = 0.3486

Ha: mean > .5
Pr(T > t) = 0.1743

Step 7: Testing the results

H0: people are not influenced by the weather



Survey's questions

Q5 - Re-formulate the ranking by imagining that you are on a day with bad weather conditions (rain, snow, extremely cold). Sort from the most used (1) to the least used (6) - drag and drop to change the order

Stata's test

One-sample t test

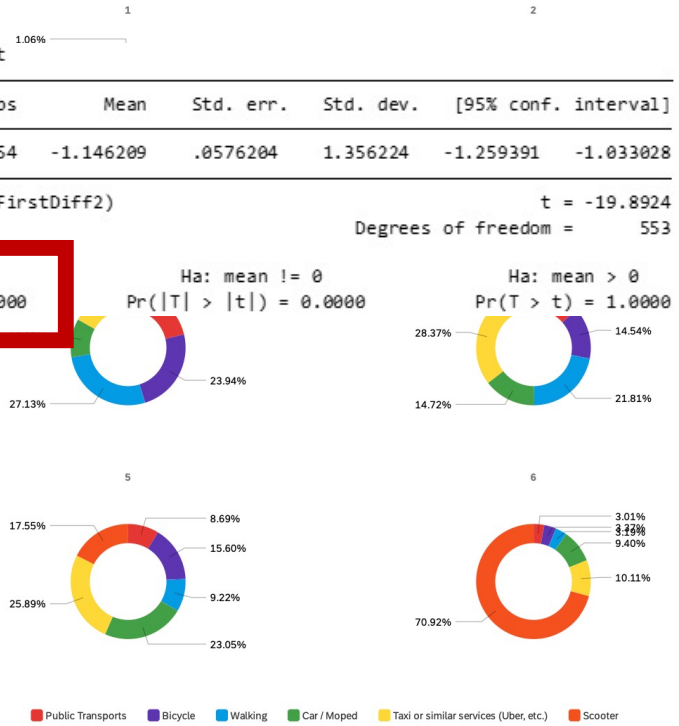
Variable	Obs	Mean	Std. err.	Std. dev.	[95% conf. interval]
FirstD~2	554	-1.146209	.0576204	1.356224	-1.259391 -1.033028

mean = mean(FirstDiff2) t = -19.8924
H0: mean = 0 Degrees of freedom = 553

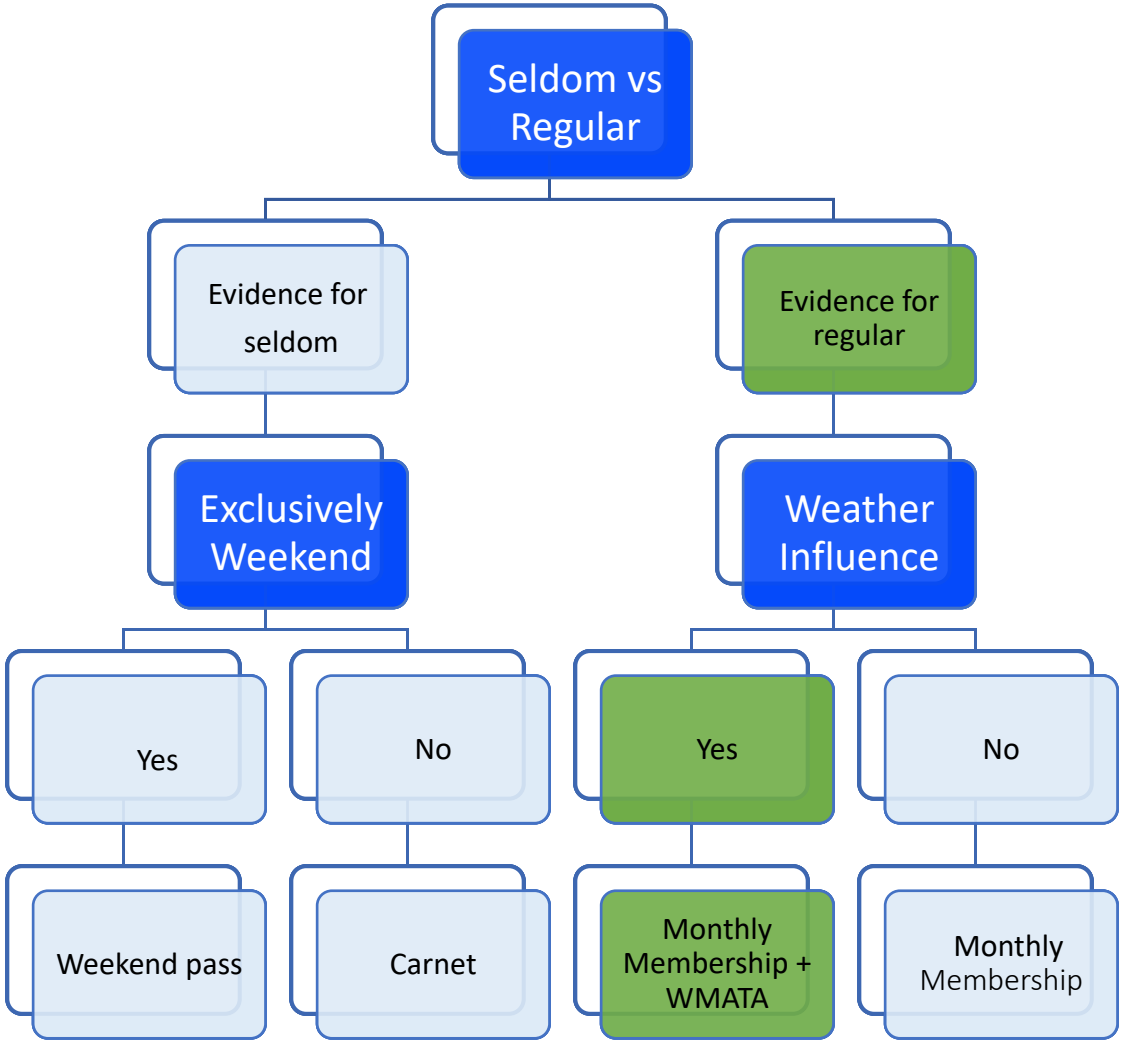
Ha: mean < 0
Pr(T < t) = 0.0000

Ha: mean != 0
Pr(|T| > |t|) = 0.0000

Ha: mean > 0
Pr(T > t) = 1.0000



Step 8: Make decisions



Monthly membership
+
WMATA

$32 / month
Unlimited 45-minute rides on a classic bike + free access to public transports

Join now

Additional tests: check true needs of regular cyclists are in line with our promotion

We want to measure how the position of public transports changes with respect to the bicycle's one in the two rankings

DiD = (average position of bicycle in the first ranking – average position of public transports in the first ranking)
–
(average position of bicycle in the second ranking – average position of public transports in the second ranking)

Stata's test

One-sample t test

Variable	Obs	Mean	Std. err.	Std. dev.	[95% conf. interval]	
DiD	554	-1.980144	.0880058	2.071411	-2.153011	-1.807278

mean = mean(DiD) t = -22.5002
H0: mean = 0 Degrees of freedom = 553

Ha: mean < 0
Pr(T < t) = 0.0000

Ha: mean != 0
Pr(|T| > |t|) = 0.0000

Ha: mean > 0
Pr(T > t) = 1.0000

More… relationship between bike sharing usage and public transports

OLOGIT public_transports bad conditions

Ordered logistic regression — Number of obs = 554; LR chi2(12) = 104.60; Prob > chi2 = 0.0000; Log likelihood = -847.70375; Pseudo R2 = 0.0581

Public_Transports_T	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
1.BikeSharing_d	-.9583053	.165018	-5.81	0.000	-1.281735	-.634876
age_encoded						
40-60	.6240827	.1835055	3.40	0.001	.2644184	.9837469
60+	.8224871	.3999759	2.06	0.040	.0385487	1.606426
Below 25	.3082643	.2966748	1.04	0.299	-.2732077	.8897362
I prefer not to specify	1.208697	1.302734	0.93	0.354	-1.344614	3.762009
gender_encoded						
I prefer not to specify	1.305462	.831013	1.57	0.116	-.3232939	2.934217
Male	.211698	.2247139	0.94	0.346	-.2287332	.6521291
Other	.1954469	.5670655	0.34	0.730	-.915981	1.306875
Profession_encoded						
Other	1.203022	.6981196	1.72	0.085	-.1652669	2.571312
Student	-.2017992	.7088695	-0.28	0.776	-1.591158	1.18756
Unemployed	-.6222749	.9200111	-0.68	0.499	-2.425463	1.180914
Worker	.2805518	.659133	0.43	0.670	-1.011325	1.572429

OLOGIT public_transports general conditions

Ordered logistic regression — Number of obs = 544; LR chi2(12) = 98.90; Prob > chi2 = 0.0000; Log likelihood = -817.38451; Pseudo R2 = 0.0570

Public_Transports	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
1.BikeSharing_d	-.8602176	.1687292	-5.10	0.000	-1.190921	-.5295146
age_encoded						
40-60	.6573211	.1920445	3.42	0.001	.2809208	1.033721
60+	.5724698	.3910592	1.46	0.143	-.1939921	1.338932
Below 25	-.0505024	.3136695	-0.16	0.872	-.6652833	.5642785
I prefer not to specify	.5767943	1.301791	0.44	0.658	-1.97467	3.128258
gender_encoded						
I prefer not to specify	2.249718	.8440637	2.67	0.008	.5953835	3.904052
Male	.2930458	.2335524	1.25	0.210	-.1647085	.7508001
Other	.2789749	.5795011	0.48	0.630	-.8568264	1.414776
Profession_encoded						
Other	2.335466	.7624699	3.06	0.002	.8410527	3.82988
Student	.9034241	.7597921	1.19	0.234	-.585741	2.392589
Unemployed	.5638125	.9733373	0.58	0.562	-1.343894	2.471518
Worker	1.308284	.7150933	1.83	0.067	-.0932731	2.709841

Those who claimed to have already used bike-sharing services had a higher probability of placing public transports in the first positions, and the marginal effect was even higher in the ranking of the bad weather conditions scenario.

Further options

Investigating more about the reasons that push people to use the service with less frequency during fall or on days with bad weather conditions

In the following weather conditions, what would motivate you the most NOT to use the bicycle?

	Causes			
	Cold hands	Risk of falling	Getting wet	Other
Rain	○	○	○	○
Snow	○	○	○	○
Low temperature	○	○	○	○

	Getting wet	All the others
Improving our promotion with additional options related to the problem of getting wet	+++	-
Do nothing/pivot	0	0

H0: getting wet is the main problem during rainy days

Stata's test

```
. ttest   Wet == 0.5 if regular == 1

One-sample t test

Variable |     Obs        Mean    Std. err.   Std. dev.   [95% conf. interval]
---------+--------------------------------------------------------------------
     Wet |     554    .5938628    .0208842    .4915546    .5528408    .6348848

    mean = mean(Wet)                                      t =   4.4945
H0: mean = 0.5                               Degrees of freedom =      553

  Ha: mean < 0.5               Ha: mean != 0.5                 Ha: mean > 0.5
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

Final Considerations

Conclusions

- Sustainable growth option with a scientific approach
- To attract casual customers towards membership
- Hypothesis supported by evidence
  - Clustering based
  - Survey based
- Monthly membership + WMATA partnership is most reliable offer

Limitations

- Riders' IDs
- Member > Casual assumption
- Survey representativeness
- Financials' weakness

Q & A

Slidesgo : https://slidesgo.com

Freepik: https://it.freepik.com