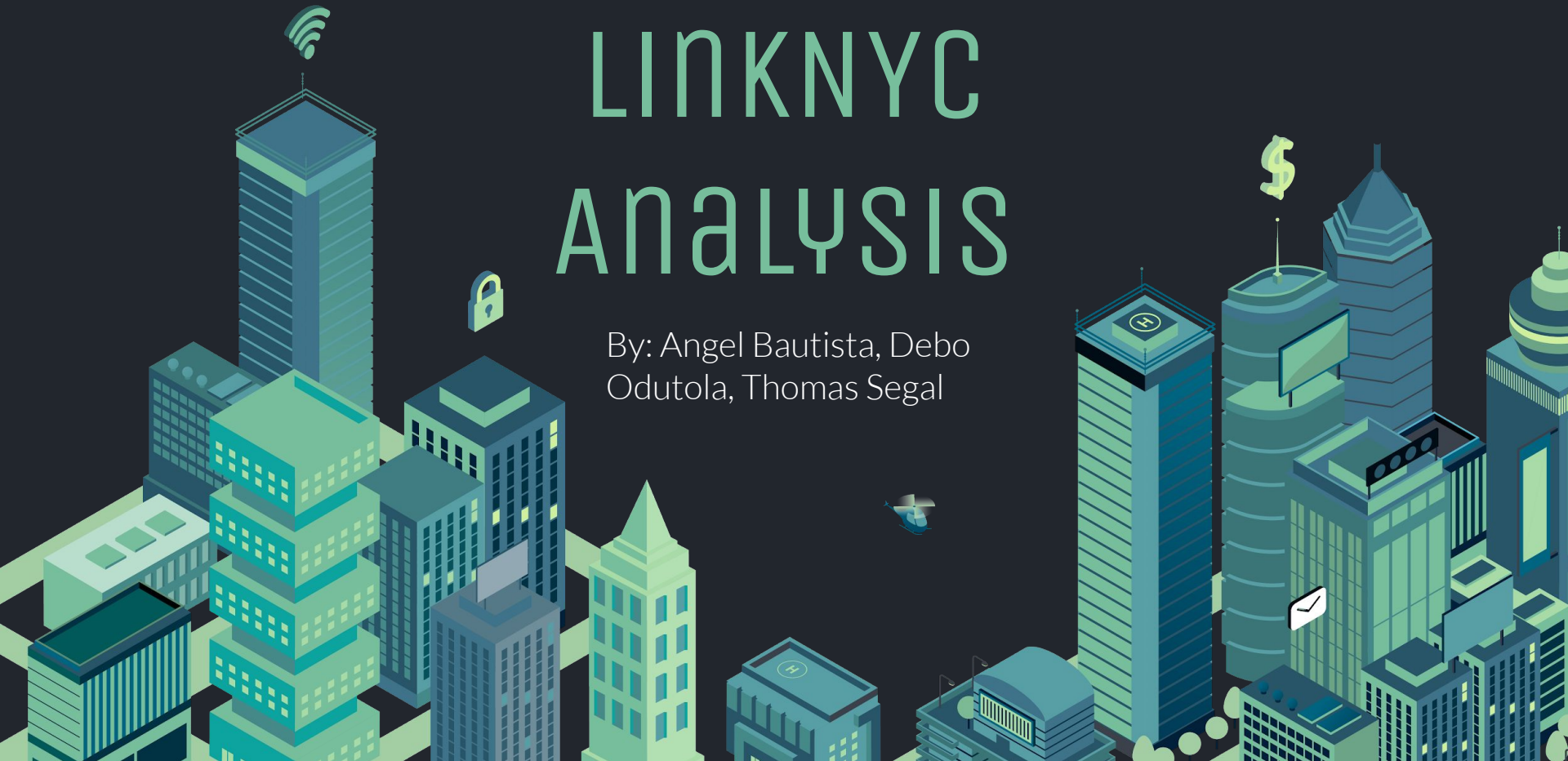
LinkNYC Analysis

By: Angel Bautista, Debo Odutola, Thomas Segal



Business Problem

LinkNYC's current kiosk distribution disproportionately serves high-traffic, higher-income areas, leaving low-income neighborhoods with limited access to free Wi-Fi and digital resources. This imbalance contributes to New York City's digital divide and undermines the program's goal of equitable connectivity. Expanding kiosk placement to underserved communities would not only promote social equity but also increase overall engagement, strengthen LinkNYC's brand value, and align with the city's digital inclusion initiatives.

AGENDA

1. EDA & KPI'S
2. Funnel
3. Features
4. Cohorts
5. RFM Segmentation

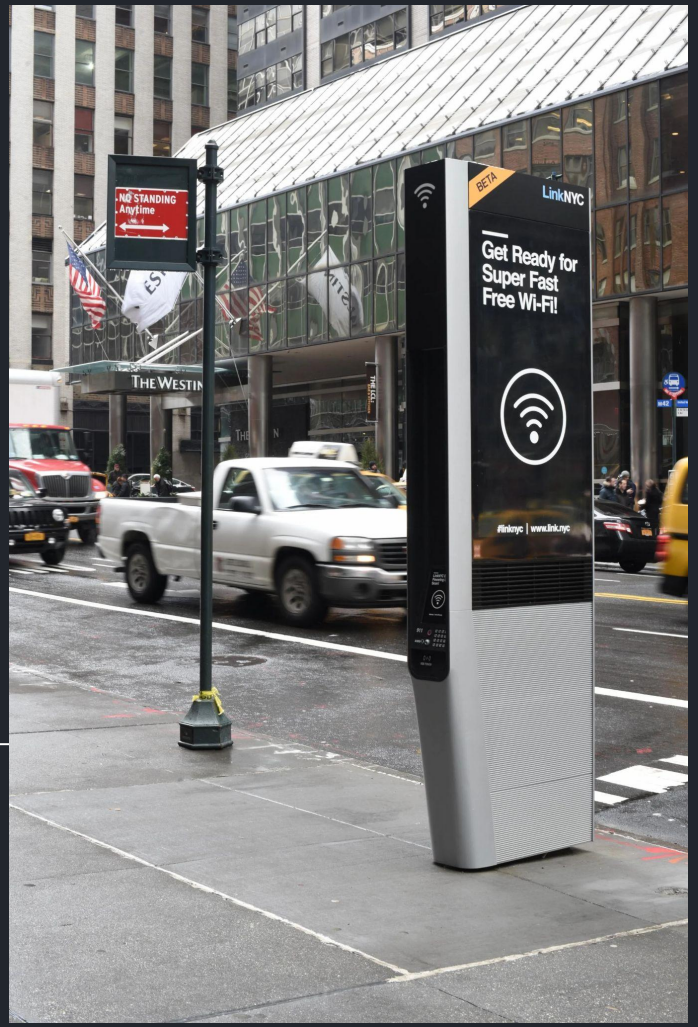
EDA & KPI

- Both datasets used in EDA (no join)
- Time units/text normalized
- Missingness checked
- Outliers checked
- Other potential data quality issues checked
- Chose activation metric
- Defined experience quality

KPI

We chose Candidate B (quality-oriented): a week is “activated” if GB per session ≥ Z (and sessions > 0).

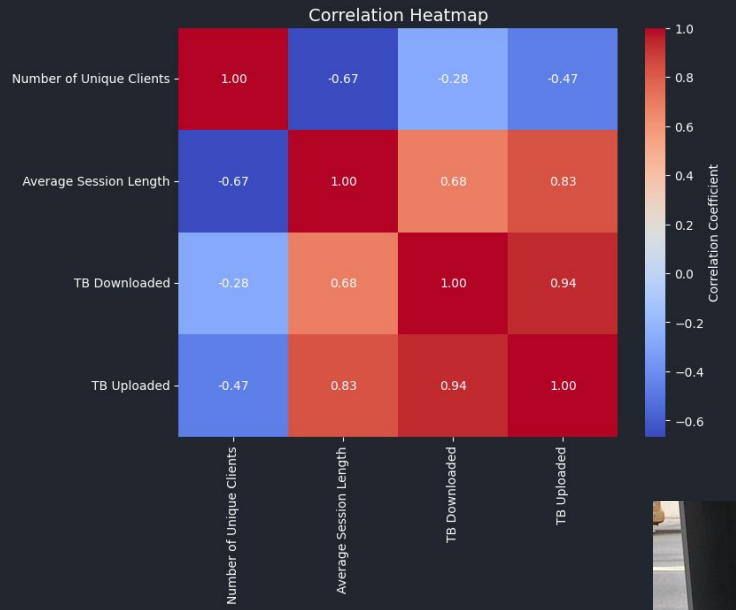
We computed the 75th percentile (Z) rather than the median to ensure activation represents clearly above-average valuable usage, which better signals a kiosk is delivering meaningful internet access.

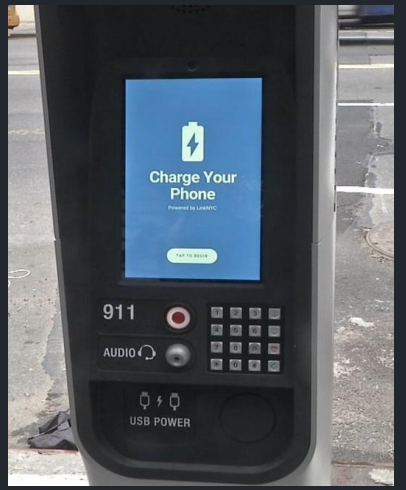


usage eda

Steps we took:

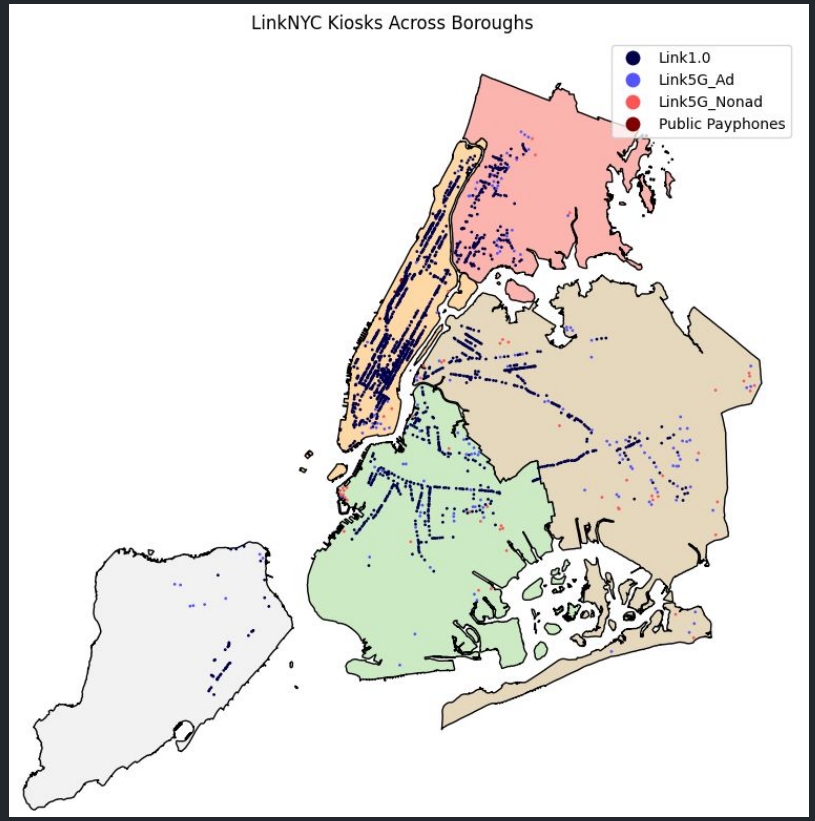
- Convert date column to datetime.
- Convert session length to minutes (float).
- Sort by week and check for missing or duplicate weeks.
- Check for outliers.
- Normalize text.





Location EDA

- Checked for duplicates
- Looked at certain fields grouped by borough
  - Installation Status
  - Planned Kiosk Type
  - Average Activation Wait (Time between Installation Date and Activation Date)



Borough	Planned Kiosk Type	
Bronx	Link1.0	186
	Link5G_Ad	26
	Link5G_Nonad	4
Brooklyn	Link1.0	299
	Link5G_Ad	36
	Link5G_Nonad	21
Manhattan	Link1.0	1206
	Link5G_Ad	17
	Public Payphones	4
	Link5G_Nonad	3
Queens	Link1.0	308
	Link5G_Ad	63
	Link5G_Nonad	27
Staten Island	Link1.0	42
	Link5G_Ad	8

Borough	Average Activation Wait
Bronx	79 days 23:46:40
Brooklyn	104 days 08:37:45.168539326
Manhattan	113 days 20:37:27.804878048
Queens	99 days 07:46:44.020100503
Staten Island	61 days 16:48:00

Borough	Installation Status	
Bronx	Live	214
	Removed Pending Reinstall	2
Brooklyn	Live	355
	Removed Pending Reinstall	1
Manhattan	Live	1224
	Removed Pending Reinstall	6
Queens	Live	394
	Removed Pending Reinstall	4
Staten Island	Live	50

2

Funnel Analysis

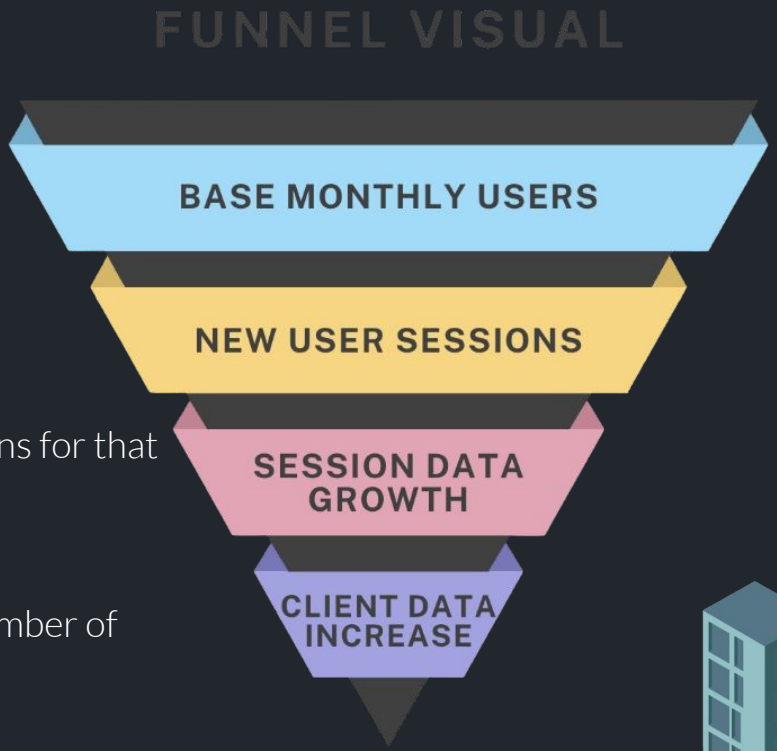
FUNNEL VISUAL

Step 1: Base amount of users for the month =1

Step 2: Increase in sessions from new users

Step 3: Increase in data transferred (GB) for sessions for that month

Step 4/ overall: increase in data transfer(GB) by number of unique clients



3

Features

Features

- 9 Features
  - 4 Essential Features
    - Total TB → Adds TB Downloaded and TB Uploaded to get a total amount of TB for the week
    - Heavy Usage Week → True if the session in the week falls above the activation threshold
    - Low Users / Long Session → True if both "Low Users", and "Long Session" are true
    - Activation Wait → Calculates the wait between when the kiosk was Installed to when the kiosk was Activated
  - 5 Assisting Features
    - TB to GB → Converts TB to GB
    - Long Session →True if the average session length for the week is ≥ 25 Minutes
    - Low Users → True if the # of users is ≤ 250,000 for the week
    - *Zoning → The zoning field uses a 3 letter code (e.g. "R2A / C1-2") to define the type of zone where the kiosk is being installed
    - Summer Months → True if the report week is in June, July, or August

Blue = Binary Feature
Red = Base Feature

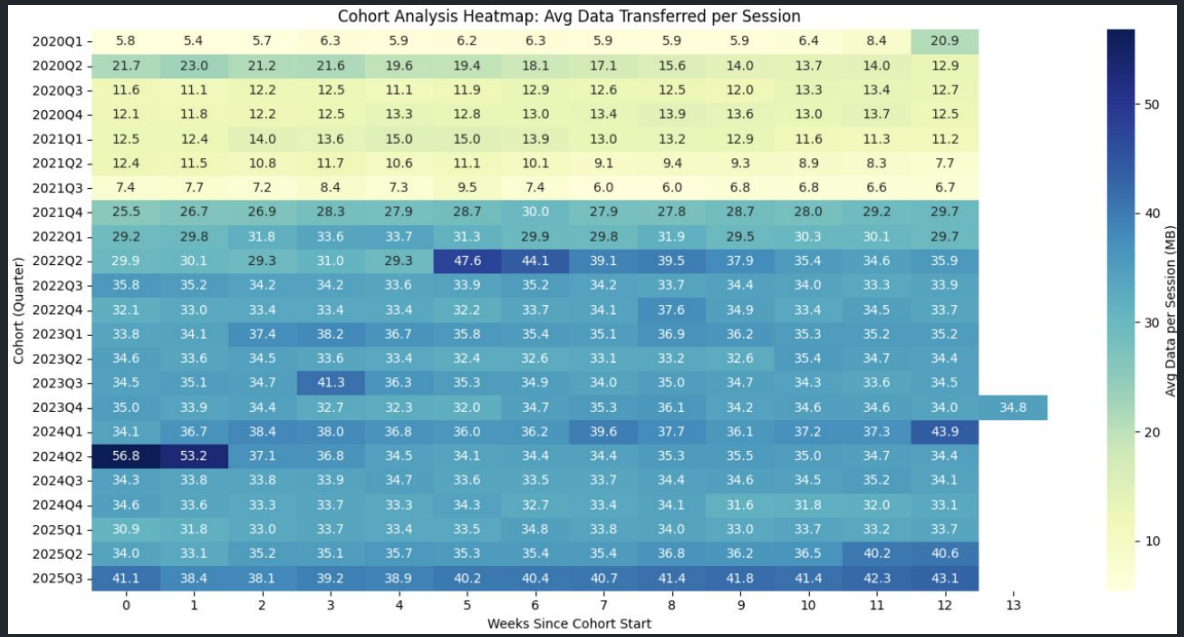
	Total TB	Total GB	heavy_usage_week	is_summer	is_long_session	is_low_users	lowuser_longsesh
290	141.31	141310.0	0	1	1	1	1
291	136.09	136090.0	0	1	1	1	1
292	142.75	142750.0	1	1	1	1	1
293	136.53	136530.0	1	1	1	1	1
294	138.75	138750.0	1	1	1	1	1
295	145.48	145480.0	1	1	1	1	1
296	143.84	143840.0	1	0	1	1	1
297	147.29	147290.0	1	0	1	1	1
298	149.70	149700.0	1	0	1	1	1
299	147.95	147950.0	1	0	1	1	1

	is_commercial	is_residential	is_manufacturing	Activation Wait
0	1	0	0	47 days
1	0	0	1	156 days
2	1	1	0	398 days
3	1	0	0	40 days
4	1	0	0	26 days

*C = Commercial Zone
R = Residential Zone
M = Manufacturing Zone
(e.g. "R2A / C1-2") = Residential Zone / Commercial Zone

4

Cohorts & Retention



Ethics/ Considerations:

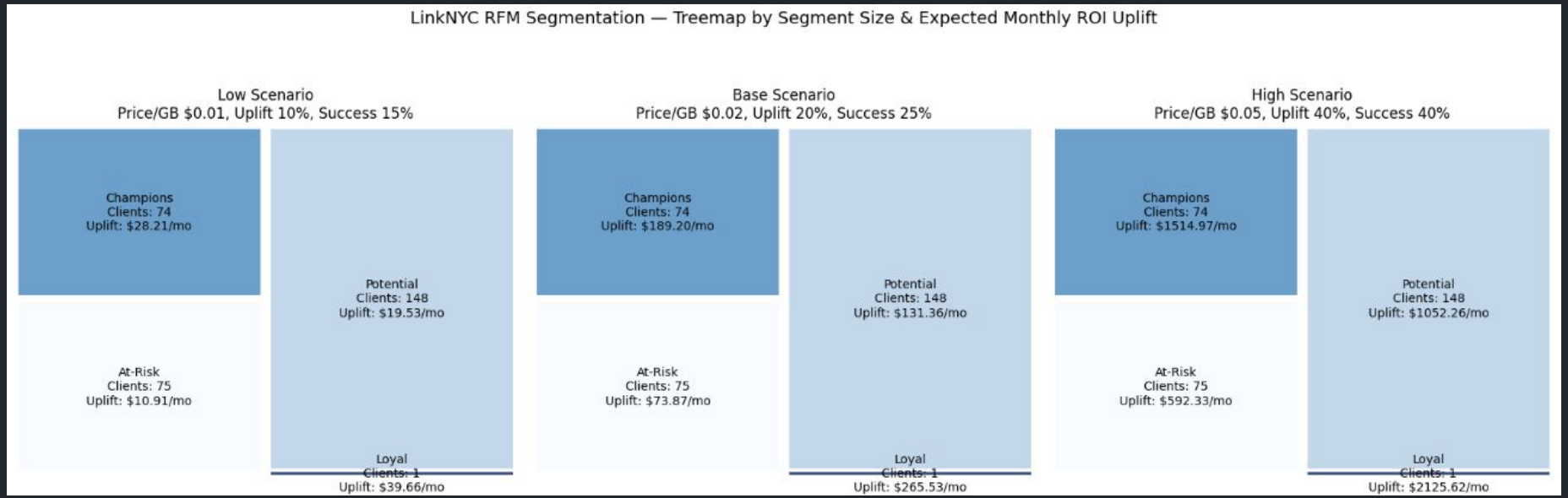
- Inherited data aggregated on weekly basis smoothes individual experiences
- Cohort performance dependant on average data transfer, instead of retention rate.

- Cohorts on quarterly basis measuring the weeks within each quarter
- Later cohorts tend to have higher average data transfer

5

RFM segmentation

RFM Treemap


LinkNYC RFM Segmentation — Treemap by Segment Size & Expected Monthly ROI Uplift
Low Scenario
Price/GB $0.01, Uplift 10%, Success 15%
Champions
Clients: 74
Uplift: $28.21/mo
Potential
Clients: 148
Uplift: $19.53/mo
At-Risk
Clients: 75
Uplift: $10.91/mo
Loyal
Clients: 1
Uplift: $39.66/mo
Base Scenario
Price/GB $0.02, Uplift 20%, Success 25%
Champions
Clients: 74
Uplift: $189.20/mo
Potential
Clients: 148
Uplift: $131.36/mo
At-Risk
Clients: 75
Uplift: $73.87/mo
Loyal
Clients: 1
Uplift: $265.53/mo
High Scenario
Price/GB $0.05, Uplift 40%, Success 40%
Champions
Clients: 74
Uplift: $1514.97/mo
Potential
Clients: 148
Uplift: $1052.26/mo
At-Risk
Clients: 75
Uplift: $592.33/mo
Loyal
Clients: 1
Uplift: $2125.62/mo

Segments explanation

Potential

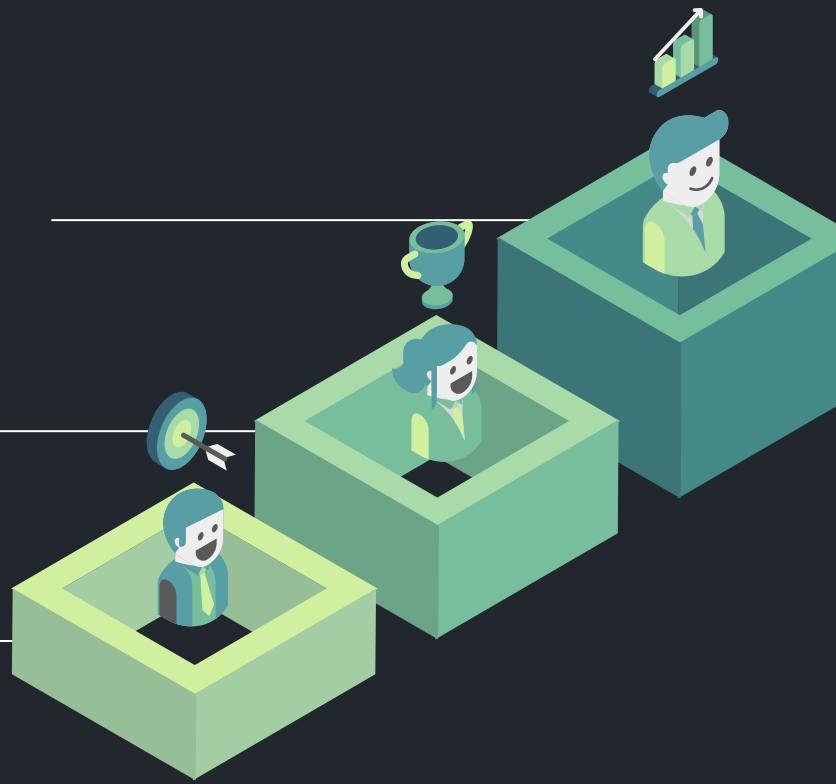
Moderate usage and recency - promising for reactivation marketing

Champions

Very recent and frequent users with high data usage - most profitable and loyal group.

At-Risk

Declining or less frequent users who could be recovered with targeted offers.



Loyal: Regular + high engagement | Hibernating: Rarely active users | Lost: Inactive for long periods

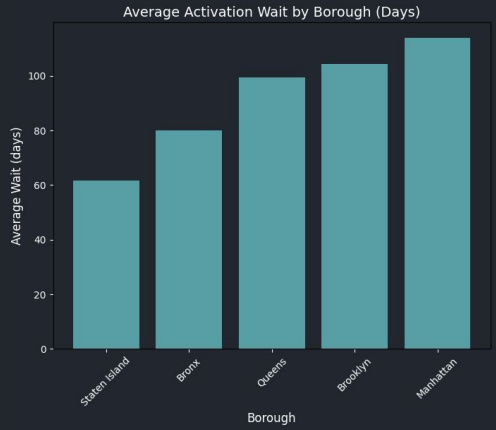
RISK & ROLLOUTS

RISKS

- Concentrating new kiosks in already connected areas could exacerbate digital divides.
- Public Wi-Fi may raise privacy concerns.
- Installations and activation take forever due to time gap between actions.

ROLLOUTS

- Use Wi-Fi density maps and socio-economic indicators to prioritize underserved ZIP codes.
- Maintain strict encryption and anonymized data logging,
- Close gap between installation and activation by increasing staff.



Thanks

Thomas

Thomasegal02@gmail.com

Debo

deboodutola@gmail.com

Angel

Bautista101905@gmail.com