Applying Iterative Design Principles to a Live Product

Name:
Date:

Step 1

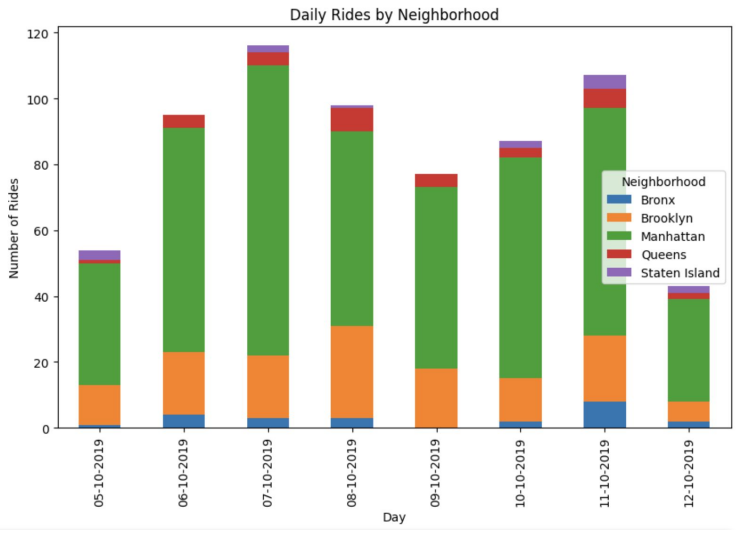
Select KPIs

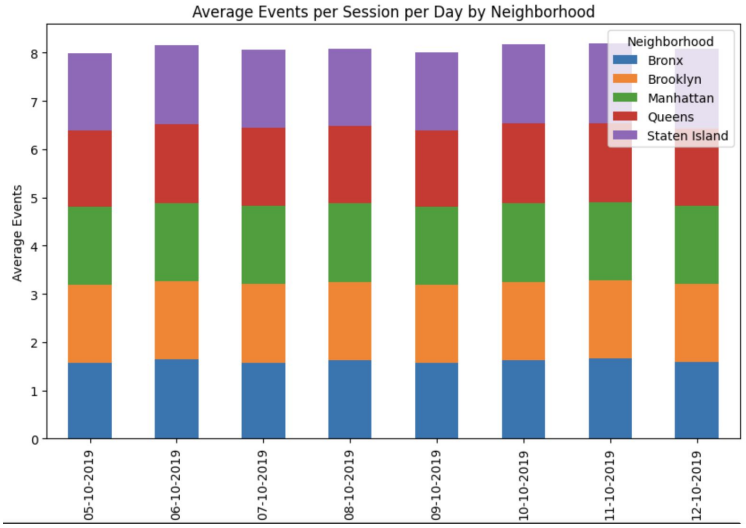
&

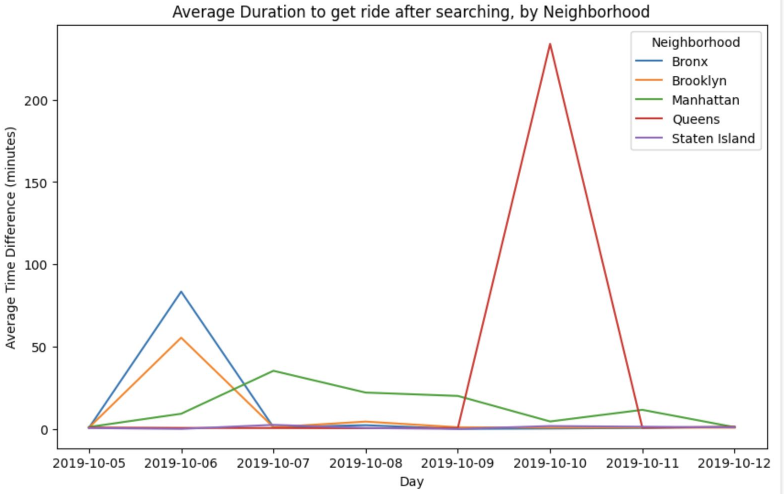
Evaluate Previous Multivariate Experiment Results

Select KPIs for Flyber Analyses

- KPIs to best match Flyber business model:
  - KPI1: Daily Rides by Neighborhood
  - KPI2: Average Events per Session by Neighborhood
  - KPI3: Average Duration, from searching to getting ride by Neighborhood
- How to calculate these KPIs
  - KPI1: Count all the events where event_type = ‘begin_ride’ per day per Neighborhood
  - KPI2: Count how many events (rows) there are grouping by session_id and day, plot the average by neighborhood over time
- Other KPIs that may be important but not calculable with current data
  - Operational KPIs that cannot be calculated are for example:
    - Duration from ordering the taxi to starting ride (no event_type defined)
    - Duration of ride itself, from begin_ride to end_ride (no event_type defined)







Describe the First Multivariate Experiment

In this multivariate test, the team is testing for two variable at the time generating 4 combinations.
The elements of change are:
1. *Tip included or excluded below the price
2. The action button “Book Flight” versus “Fly Now”

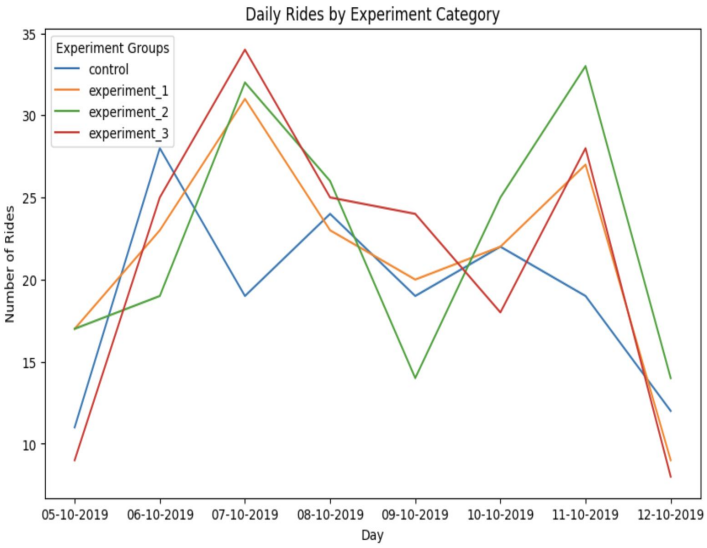
Control

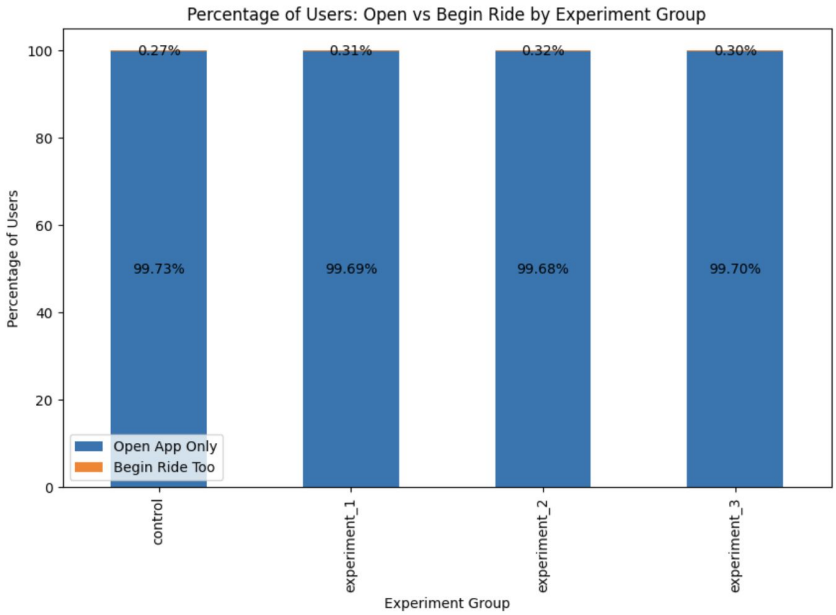
Experiment 1

Experiment 2

Experiment 3

Review Multivariate Test Results: Visualization





Experiment categories 2 and 3 seems to have performed overall better. These two have in common the omission of the '*tip included' indication. Control performs overall worse than all but 3 is not always better than 2, hence excluding the assumption that the action button 'Fly Now' is better than 'Book Flight'.

Looking at the second graph we can see that in % experiment 2 has converted more users. This is the no tip indication and 'Book Flight'. Let's see if the results are significant next.

Experiment 1 - not significant

	Visitors	Conversions	Conversion Rate
A	56236	154	0.27%
B	56218	172	0.31%

Confidence

90% 95% 99%

CALCULATE

Result not significant!

Variant B's conversion rate (0.31)% was higher than variant A's conversion rate (0.27)%, but you cannot say, with 95% confidence, that variant B will perform better than variant A.

p value

`0.1583`

Experiment 2 - not significant

	Visitors	Conversions	Conversion Rate
A	56236	154	0.27%
B	56508	180	0.32%

Confidence

90% 95% 99%

CALCULATE

Result not significant!

Variant B's conversion rate (0.32)% was higher than variant A's conversion rate (0.27)%, but you cannot say, with 95% confidence, that variant B will perform better than variant A.

p value

`0.0837`

Experiment 3 - not significant

	Visitors	Conversions	Conversion Rate
A	56236	154	0.27%
B	56516	171	0.30%

Confidence

90% 95% 99%

CALCULATE

Result not significant!

Variant B's conversion rate (0.30)% was higher than variant A's conversion rate (0.27)%, but you cannot say, with 95% confidence, that variant B will perform better than variant A.

p value

`0.1842`

Review Multivariate Test Results: Significance Test

Determine if there was a significant difference between the experiments and control states.

- Explain how you would perform a t-test to determine if the experimental results had a greater impact on the booking conversion rate than the control state
  - the t-test helps assess whether the observed difference between two groups is likely a real effect or just due to random variation in the samples
  - Formula: t = mean1 - mean2 / sqr ((stdev1/sample size1)^2 + (stdev2/sample size2)^2)
  - We can replace e.g. “Control” with 1 and “Experiment1” with 2 in the formula above
- List the test results (p value) for each experiment compared to the control
  - Pictures slide before
- Using the statistical significance calculator of your choice, determine which experiments, if any, had a significant result at the 95% level. Include your calculations as part of your explanation
  - Inserting the values in the online calculator, no experiment variation is below p value of 0.05 (the closest, 32% reaches ~0.08)
- Based on your statistical significance calculations, recommend if any of the experiments should be expanded
  - Given that all of the experiments in this multivariate **do not show significant results**, if we deem the volumes reached sufficient, we can conclude that the two variables have no effect on the user to book a ride, **we should discontinue** this test altogether

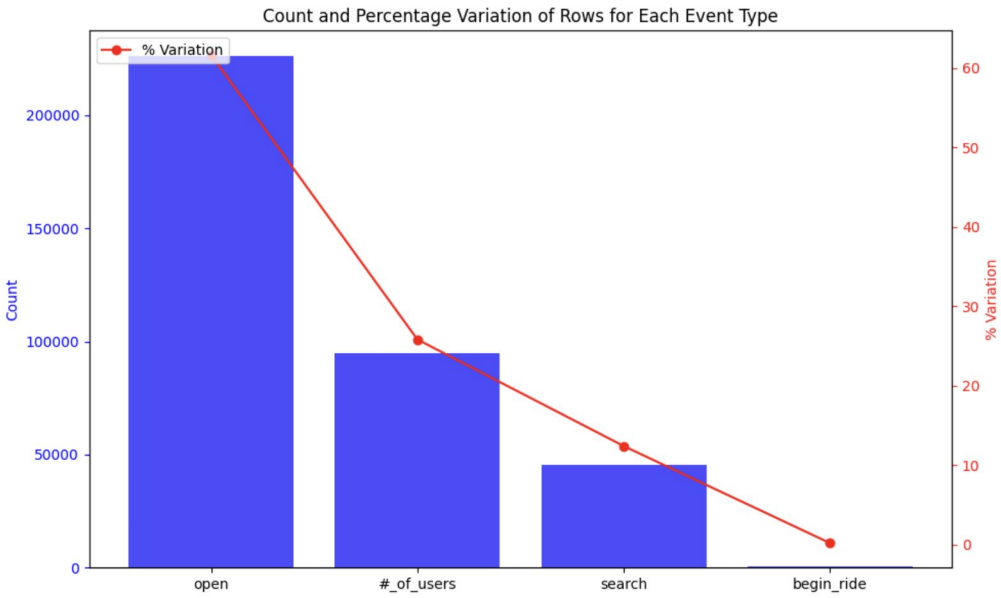
Step 2

Funnel & Cohort Analyses

User Funnel

Identifying the different stages the user funnel

- Based on the event types in the data provided, list the 3 or more steps a user can take from opening the app to final booking of a ride
  - **Open, #_of_users, search, and begin_ride**
- Provide a graph showing the funnel from step to step, including drop off rates.
  -




```
open           226155
#_of_users      94748
search          45503
begin_ride        677
Name: event_type, dtype: int64
open           61.608683
#_of_users     25.811056
search         12.395834
begin_ride      0.184427
Name: event_type, dtype: float64
```

User Segments

- Identify 2 demographic attributes present in the data that allow for segment analysis
  - Age and neighborhood
- For each demographic attribute, provide the number of users in each segment group
- For each demographic attribute, identify the segment group with the largest number of users

age	user_neighborhood	count
50+	Manhattan	123645
40-49	Manhattan	66599
18-29	Manhattan	40165
50+	Brooklyn	35379
30-39	Manhattan	26850
40-49	Brooklyn	19183
18-29	Brooklyn	11641
50+	Queens	8603
30-39	Brooklyn	7677
50+	Bronx	5190
40-49	Queens	4720
50+	Staten Island	3378
18-29	Queens	2835
40-49	Bronx	2740
30-39	Queens	1930
40-49	Staten Island	1926
18-29	Bronx	1686
30-39	Bronx	1186
18-29	Staten Island	1037
30-39	Staten Island	713

Segment Analysis of Funnel

Identify Opportunities for Improvement

- Perform a funnel analysis by segment for all identified demographic attributes and describe the results
  - For age, 18-29 converts the most to search, but 40-49 converts overall the most to begin ride
  - For Neighborhood, Queens converts the most to search and bronx converts the most to begin ride
  - Combinations of the demo attributes, Bronx 40-49 converts the most to rides
- If underperformance for a segment in an attribute is identified, add a visual showing the average funnel conversion by segment group for that demographic
  - Manhattan Boomers converts the worst to begin rides, plot below

Instead of analysing the % change between 2 bins at the time, I show the overall % distribution. Same conclusions can be derived; the distribution of the 50+ group is skewed to the "left" towards open and #_users more than all the others ages, showing how they converted less to search+begin_ride

Percentage Variation Table for Age:

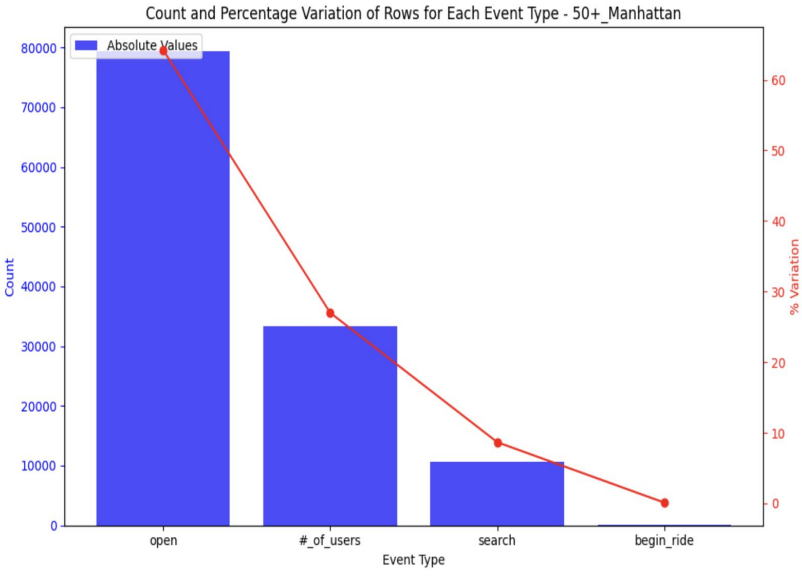
event_type	#_of_users	begin_ride	open	search
age				
18-29	24.792553	0.219650	59.057946	15.929851
30-39	24.707999	0.260715	59.338826	15.692460
40-49	24.762525	0.268998	59.112307	15.856170
50+	26.949119	0.110673	64.281620	8.658588

Percentage Variation Table for User Neighborhood:

event_type	#_of_users	begin_ride	open	search
user_neighborhood				
Bronx	25.532309	0.212924	61.960748	12.294020
Brooklyn	25.797239	0.182729	61.698701	12.321332
Manhattan	25.838163	0.184250	61.558974	12.418613
Queens	25.586024	0.171384	61.764706	12.477886
Staten Island	25.971080	0.198469	61.539552	12.290899

Percentage Variation Table:

event_type		#_of_users	begin_ride	open	search
age	user_neighborhood				
18-29	Bronx	24.495848	0.177936	59.489917	15.836299
	Brooklyn	24.937720	0.171807	59.221716	15.668757
	Manhattan	24.757874	0.236524	59.026516	15.979086
	Queens	25.008818	0.176367	58.095238	16.719577
	Staten Island	24.397300	0.289296	60.366442	14.946962
30-39	Bronx	24.030354	0.337268	60.286678	15.345700
	Brooklyn	24.749251	0.247493	59.749902	15.253354
	Manhattan	24.748603	0.256983	59.188082	15.806331
	Queens	24.145078	0.310881	59.378238	16.165803
	Staten Island	25.385694	0.280505	58.906031	15.427770
40-49	Bronx	24.416058	0.364964	59.525547	15.693431
	Brooklyn	24.709378	0.265860	59.224313	15.800448
	Manhattan	24.778150	0.270274	59.071458	15.880118
	Queens	24.618644	0.211864	59.258475	15.911017
	Staten Island	25.597092	0.259605	58.463136	15.680166
50+	Bronx	26.801541	0.115607	64.431599	8.651252
	Brooklyn	26.897312	0.127194	64.278244	8.697250
	Manhattan	26.996644	0.105140	64.236322	8.661895
	Queens	26.630245	0.116239	64.884343	8.369174
	Staten Island	26.791001	0.118413	64.209591	8.880995



Step 3

Hypothesis & Next Steps

Review Qualitative Data

- Read user interviews to understand “why” any funnel under-performance seen in Step 2 might occur
  - Most “unsatisfied” users seem to cite the main reason being that Flyber cannot be booked as easily as a taxi, or using voice calls
  - Knowing also from the data that the 50+ age groups is the one converting suboptimally, we could include tailored features to /B-test specifically to this age group

- List your hypothesis for what customer need is being under-served
  - The underserved customer need is velocity of booking, user-friendliness of booking

- Provide 3 or more quotes as evidence for this hypothesis
  - I just hail a taxi or tell my phone to call a cab to go to a certain address (I'm always on the phone
  - I have a personal car service on call. My assistant books Flyber whenever I'd be travelling during peak NYC traffic hours. Time is money and Flyber saves me time! But I let my assistant actually book the Flyber because the first few times I tried booking, the instructions were too small.
  - Before Flyber, I'd call a taxi service on the phone.

Suggested Features & Experimentation Plan

- We believe funnel drop at 50+% after opening the app to be because the app is not as friendly to use (to the 50+ age group specifically), and it cannot be used by phone calls. And that by Improving the A/B testing phase for Searching the Flyber we will see a 20% increase in customer getting to the begin ride page.

- Suggest 2 or more features that would match your hypothesis and determine a plan for multivariate testing, including describing the control and experimental conditions
  - This hypothesis can be observed as older customer segments decrease more in the funnel. Customers from busiest neighborhood like Manhattan also have a higher bounce rate, indicating a need to ease the booking procedure
  - Multivariate testing can be conducted at the first landing page:
    - 2 variables -
      - **Feature 1: reserve the most convenient Flyber** (combo between cheap/close to me) **option enabled/disabled**
      - **Feature 2**: Include **option to reserve by calling (big font size)** at opening page
    - The multivariate test will then include 4 experimental group including Control. Specifically the roll out will include:
      - Control will remain as it is, Feature 1 and 2 not enabled
      - Group 1: Feature 1 and Feature 2 simultaneously enabled
      - Group 2: Feature 1 is enabled and Feature 2 is disabled
      - Group 3: Feature 1 is disabled and Feature 2 is enabled
    - This would tackle most of the issues encountered in the interview dataset

Multivariate Test	Feature 1 - Quick Book	Feature 2 - Call Flyber
Control	Disabled ❌	Disabled ❌
Group 1	Enabled ✅	Disabled ❌
Group 2	Disabled ❌	Enabled ✅
Group 3	Enabled ✅	Enabled ✅

Suggested Features & Experimentation Plan

- Determine who should be exposed to the experimental changes
  - Only the 50+ age group should be exposed to the changes, hence only expecting positive impact. If the features were rolled out to all users, or more age groups the expected impact could also be negative, overall defeating the purpose of the intervention
  - Once chosen the confidence interval, one should choose a large enough sample size to justify the results. Theoretically the population size can be ignored in the math but empirically it does give a sense of meaning to a business expert.
  - Alternatively the whole population can be partitioned randomically into 4 buckets, 1 control and 3 experiments
  - The

- List any additional metrics that would be helpful to collect from your suggested features
  - Time spent in the open page before bounce off & Time spent after having clicked the new buttons closest/cheapest to scout for any trends
  - KPIs that can be build are many