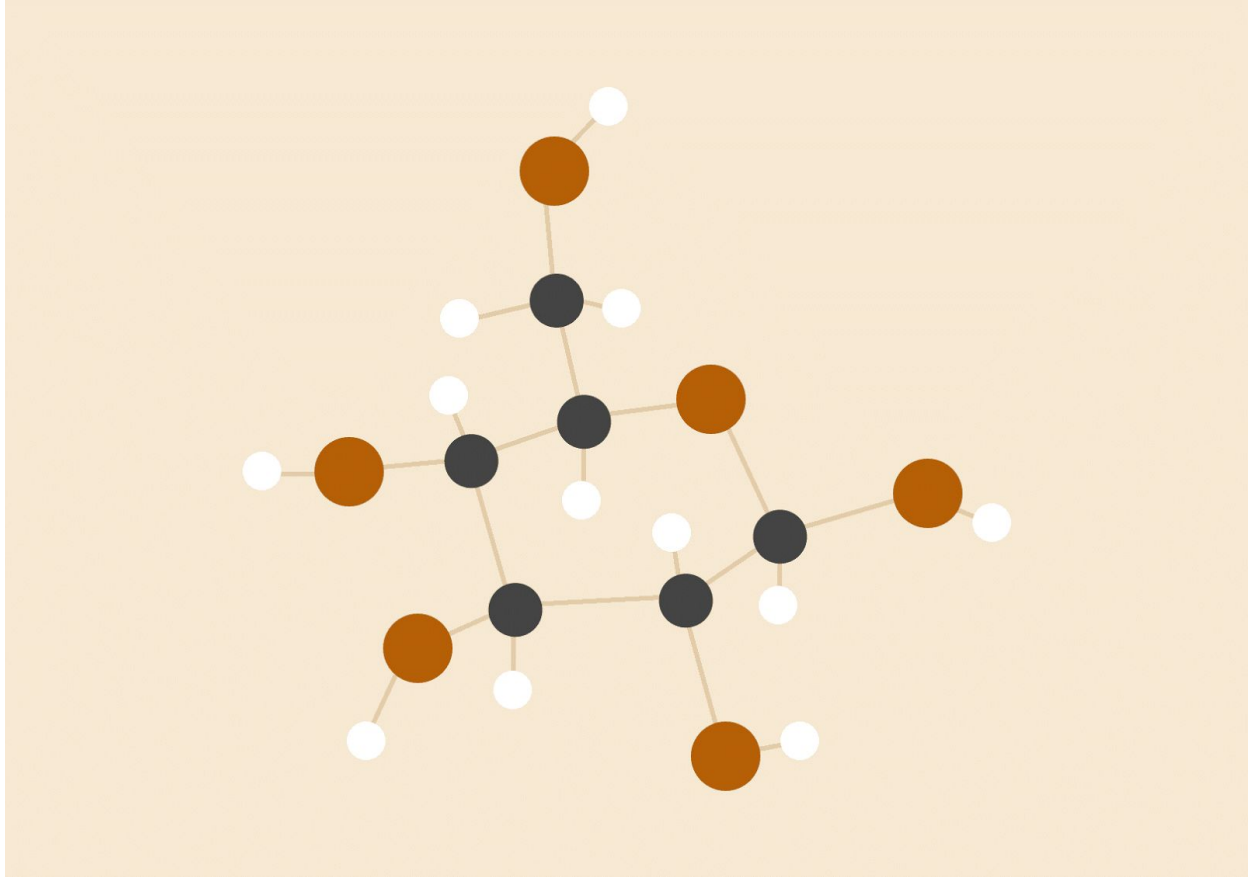# Team 25 Datathon REPORT

Luke Roantree, Xingxin Lu,

08.04.2017

Topic question

Question:

Our team set out to optimise Uber's marketing campaigns based on which age groups are their main users in different areas.

In this, we assume they have a campaign that covers all of New York that targets their main users overall (around 20-30 year olds) and we recommend they expand this to include many smaller targeted local campaigns which focus on the difference from the NY average in terms of usage by different age brackets in individual areas.

For example, in an area of NY there it is likely that the main usage of Uber will be in the 20-30 year old age group; but these people are already targeted by the large campaign and if there are 50% more users in the 40-44 year old age group in this local area than the 40-44 year old group for the NY average, it would be very useful to know this and target them in the local campaign.

Dataset uber_trips_2015, demographics and zones were used to display the findings.

Non-Technical Executive Summary

Key findings

Based on distribution for all of New york, Uber users tend to be in the xx, with very few in yy, etc. different areas however tend to have slight variations from this distribution. We believe it could be useful for Uber's adverting campains if they had an overall campain targing xx bracket mostly, bt also local compains based on the different of local

Technical Executive Summary

Methodology and approach

uber_trips_2015 → Zones → demographics

pickup_location.ie   nta_code   age_brackets
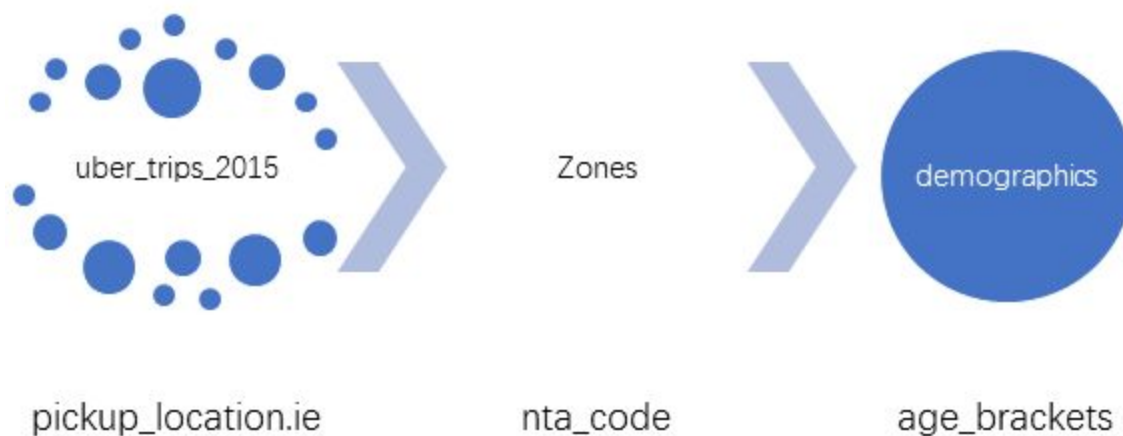
The Zones and Demographics tables were merged by their nta_code values, and the resulting table was merged with the Uber_trips_2015 using their pickup_location_id codes. Based on findings on both datasets, age_brackets(14 total) was researched in demographics dataset, and final results that included three datasets show the importances and need for optimisation of Uber's marketing campaigns based on which age groups are their main users in different areas.

RESULTS

The chart below shows the probabilities of an uber passenger being in a given age category (divided up in groups of 5 year intervals). As expected, most users are around 20-24 and 25-29. The spike at the far end is to be expected as this covers the entire 65+ category. We use this distribution as a baseline to compare the other localised distributions to.
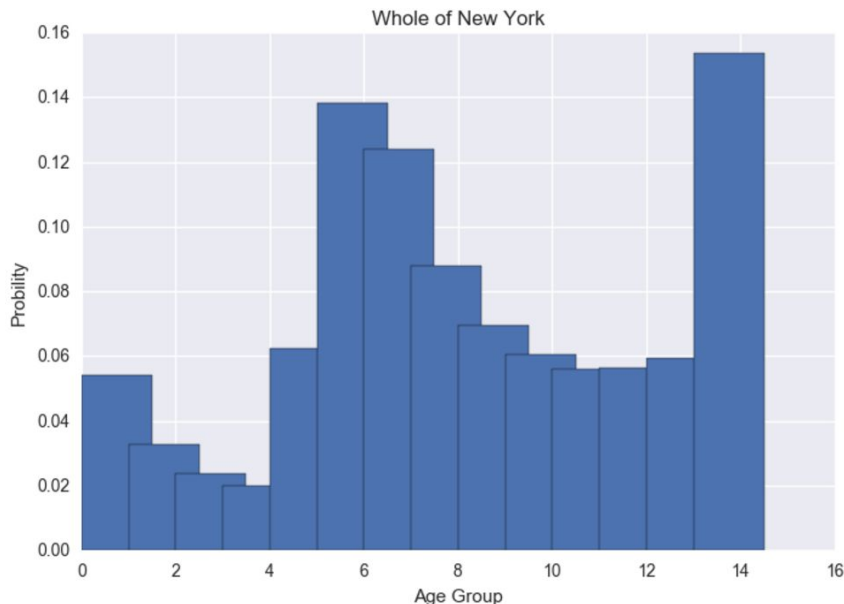
This next chart shows an example of the distribution for one zone, notice how it differs slightly from Whole New York

Lenox Hill West
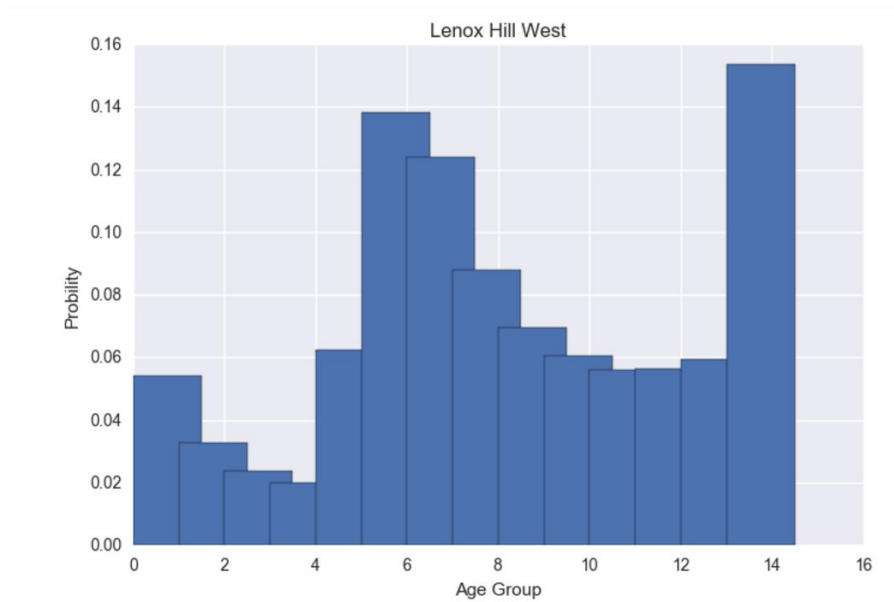
The below chart shows the differences in probability; implying that in this area there is more in the 30-34 category and uber should focus there