1. Number of active users in a given period of time. Use of advanced features beyond sign in.
2. Average time spent using new app. Average number of interactions.
3. App rating in the app store.
4. Changes in Earnings between old and new app per driver per trip or total.
5. Idle time between trips. Looking for customers.
6. Changes in ratings pf the driver.

# Success metrics

* Primary success metric should be active user count.  
  Active user is defined as any partner that frequently uses advanced features of the app. For example: Use of heat map, action on the Uber feed, use of earnings and ratings tabs.  
  Success could be defined as higher percentage of active users compared to the old app.
* Secondary metrics can be:
  + - * 1. App rating in the app store
        2. Significant improvements in partner’s average earnings per trip
        3. Significant improvement in partner’s rating

# Testing

1. For regions where both old and new apps are being used:
   1. Calculate fractions of active users for both apps
   2. Calculate average earnings per trip for both apps, across all partners
   3. Calculate average ratings of partners for both apps, across all partners
2. For people that switched to new app:
   1. Calculate how many were ‘active’ before and after (Testing whether the new app is more engaging)
   2. Calculate if the switch to a new app caused a statistically significant increase or decrease on average earning per trip.
   3. Calculate if the switch to a new app caused a statistically significant increase or decrease on the rating of the partner.
3. Compare apps store ratings for comparable periods

First test, comparing app performance across all partners can yield statistically significant results faster, since we will be looking at all partners and average across all.

Second test will be more precise measurement but will need longer time since calculation will be made on individual partner level.

Third test is straightforward and can yield results relatively fast.

Based on the result(s) of the first test (and probably third), we can judge on the duration or implementation of the second test.

# Recommendation

Based on the defined metrics (% of active users, app rating, earnings per trip, rating of the partner):

* if the new app performs better in at least one metric, especially primary, while other metrics are unaffected, we would recommend launch of the new app.
* If the new app is performing worse in at least one metric, especially primary, while other metrics are unaffected, we would recommend to roll it back.
* In all other cases we would recommend further studies if feasible.