**Problem Set 5 - Uber**

**BA830**

**Due Date: April 26, Noon**

In this assignment, you will conduct a data analysis in order to **inform Uber’s decision of whether to change the Uber Express Pool wait time from 2 minutes to 5 minutes.** You are given a dataset with an experiment using a switchback design (which varies the waiting period across cities, days, and times of the day). Be prepared to discuss this case, in the style of an HBS case.

**Note 1:** Please carefully read the Uber from your coursepack. Pages 9 to 13 are especially relevant to this analysis but you should read all of it.

**Note 2:** Exhibit 11 contains the variable definitions, make sure to read them in order to understand the data. Note that your dataset contains one other variable, which is the revenue that Uber takes in from passengers.

**A: Summary statistics about the data.**

1. Read the HBS Case. What is the difference between Uber POOL and Express POOL? No more than two sentences. (10 points).
2. How did Uber use surveys in designing Uber Express Pool?

No more than two sentences. (10 points).

1. Suppose Uber was considering a new algorithm to recommend ride destinations in the app. Which type of strategy should they use (A/B Test, Switchback, Synthetic Control)? No more than two sentences. (10 points).
2. Suppose Uber was considering a radio advertising campaign. Which type of strategy should they use (A/B Test, Switchback, Synthetic Control)? No more than two sentences. (10 points).
3. Create two new columns in the dataset that represent the total number of trips for both pool products and the profit from these products. (10 points)
4. Plot the average number of trips as a function of the time of the day. Describe a reason why this pattern exists. (20 points)

**B: Regression analysis of outcomes.**

1. Conduct a regression analysis of the experiment (considering the outcomes: revenue, total\_driver\_payout\_sr, rider\_cancellations, total\_trips). Make sure to think carefully about the correct regression specification. The regression output should be easy to read, so use ‘etable’ or ‘modelsummary’. What do you learn in words from this regression analysis (no more than 5 sentences but it can be less)? (30 points)  
     
   **Hint:** We should control for the fact that different times of the day and different days have different demand patterns. (Please refer to p.13 of the HBS article to see why)
2. One of your data scientists suggests that the optimal wait time may differ by whether it’s a commuting period. Test whether the effects of a 5 minute wait period on total trips and cancelations differ by whether it’s a commuting period. Which policy works better during commute times? (10 points)

**C: Be prepared to discuss the following questions during class.**

* What should Uber do next? This opinion should be based on the evidence above.
* Who is Uber’s customer?
* What is Uber’s value proposition?
* How do Express Pool, Pool, and uberX differ?
* Why did Boston have a lower adoption of Express Pool?
* What is the role of Trip Parties?
* Who are the key stakeholders in the decision to launch the 5 minute wait time?
* What are the advantages and disadvantages of the 5 minute wait time?