# Part 1: Supply Data Analysis

Important part of our business is a supply/demand balance. We can’t control demand but we can shift some supply to necessary hours to cover more demand during peaks.

As part of the task you will have sample **supply** and **demand** data over a few weeks in a single city a few weeks after launch.

**We need to understand:**

1. What is the supply to demand dynamic and whether they match? (when are the discrepancies?)
2. Where are the hours of oversupply? Can we shift some of them to undersupply hours?

**Needed output:**

1. Show which 36 hours in a week are most undersupplied. Show/describe your decision based on sample data.
2. 24-hour curve of average supply and demand (to illustrate match/mismatch).
3. Visualisation of hours where we lack supply during a weekly period. This one we can send to drivers to show when to online for extra hours.
4. Estimate number of hours needed to ensure we have a high Coverage Ratio during most peak hours.
5. Calculate levels of guaranteed hourly earnings we can offer to drivers during 36 weekly hours with highest demand without losing money + how much extra hours we want to get to capture missed demand.
   1. Assume that Finished Rides have an average value of €10 (80% goes to driver, 20% is our revenue).
   2. Assume the same level of demand with increased supply, base it on RPH over 3 hour periods, but with increased supply.
   3. Assume that with extra hours we will capture “missed coverage” or people attributed to “People saw 0 cars” in demand data.

### [Sample Supply data info (Hourly DriverActivity.csv)](https://docs.google.com/spreadsheets/d/1B90TAxLOSQj5yZnstew6iHTOicd5PE7qUzx3sfBNT8o/edit?usp=sharing)

Shows aggregated driver activity in a city for a given period.

**Details:**

* Hourly data for 5 full weeks from 2016-11-14 until 2016-12-18
* Real data from a recent launch in a competitive city (2 big apps for years)
* Fields:
  + **Date** – date + hour for which the row of data is presented
  + **Active drivers** – number of active drivers (any level of activity) available during time period
  + **Online (h)** – total supply hours that were available during time period
  + **Has booking (h)** – total hours during which drivers had a client booking (any state)
  + **Waiting for booking (h)** – total hours which drivers spent waiting for booking
  + **Busy (h)** – total hours which drivers were not available to take orders in
  + **Hours per active driver** – average number of hours each driver was online during time period
  + **Rides per online hour** – aka **RPH** – avg. finished trips per online hour during period
  + **Finished Rides** – number of finished trips during period
* Note the data is sorted with more recent data first
* Note that is time period has 0 values in all columns, it would be skipped (no row)

### [Sample Demand data info (Hourly OverviewSearch.csv)](https://docs.google.com/spreadsheets/d/1f8BdyMAw7PVs1p0Vz-DAQoB77JV5TNqTppsP_IAOgMo/edit?usp=sharing)

Shows how many people saw a car in the app when setting the pickup marker on the map. If you saw a car at one point and did not see a car later, you are counted in both columns in that period.

**Details:**

* The data is from the same period as Supply data above.
* Fields:
  + **Date** – date + hour for which the row of data is presented
  + **NB:**
  + **People saw +1 cars (unique)** – number of users who saw a car.
  + **Coverage Ratio (unique)** – % of users who saw the car.
* Note the data is sorted with more recent data first
* Note that is time period has 0 values in all columns, it would be skipped (no row)