Questions to answer:

- What is the supply to demand dynamic and whether they match?
- Where are the hours of oversupply? Can we shift some of them to undersupply hours?

Analysis

It looks that the trickiest part of the task is how you define the supply and demand from the data.

While exploring the data the following dependencies were discovered:

- Hours per active driver= Online (h)/ Active drivers (as in description)
- Rides per online hour = Finished Rides/ Online (h)
- Coverage Ratio (unique)='People saw +1 cars (unique)'/('People saw 0 cars (unique)' + 'People saw +1 cars (unique)')*100
- Online (h) = Has booking (h)+ Waiting for booking (h) (in the table displayed rounded so doesn't match perfectly)
- Demand = People saw 0 cars (unique) + People saw +1 cars (unique)

Typical driver activity consists of:

- Waiting for booking (staying at the parking or cruising)
- Driving to pick-up a client (pick-up destination)
- Driving with client
- Dropping off the client (drop-off destination)
- Time of inactivity
- Time of the ride = time to pick-up the client (waiting time for a client) + driving time

Exploration

To explain the way I proceed with the exploration, let's take and analyze the first row of the data:

There are 52 active drivers and in an ideal state there will be 52 active hours online (available). But, in fact, in average the driver was online (active for booking) only for 18 minutes during this hour. So it can happen that all the active drivers were active during the first half of an hour but all the demand could appear in the second half. In fact, from the data we can see that this guess is luckily not true.

The actual **match between demand and supply can be defined by the field 'People saw +1 cars (unique)'** and, in fact, 32 customers were able to order a ride.

There is no information about "started rides" within an hour and only the number of 'Finished rides' is provided. Some of this 'Finished rides' may have been started in previous hour(s) and are covering the demand from previous period.

The "**lack of supply" can be defined by the number of 'People saw 0 cars (unique)' or with (100-'Coverage Ratio')/100 in relative numbers**. Despite the fact that there were 52 Active drivers and the **total (possible gross) demand** was only 41 customers (who tried to order a ride and saw +0 cars=32+9) there was undersupply of 9 possible rides.

So the problem was not due the number of drivers, but due to the fact that some of the drivers were not online at the time of demand.

From the other view, data also tells us that during the hour the number of 'Finished rides' was only 12, this number provides the **factual demand and actual match,** which is only 12/32=37.5% of those who saw the car and 12/42=29.3% from all customers that

were checking the application for availability and price of the ride. So **only 12 rides out of the 41 possible rides (potential gross demand) and with 52 active drivers generated money flow for the company (were billed)**.

If we add 'Busy(h)' to 'Online(h)' it will still makes only 56% (29h) of the ideal 52h situation. There is no information provided in the data about rest of the time - when drivers were both not 'Online' and not 'Busy' – and we can suppose it is due to the fact that some drivers have finished or not started the shift.

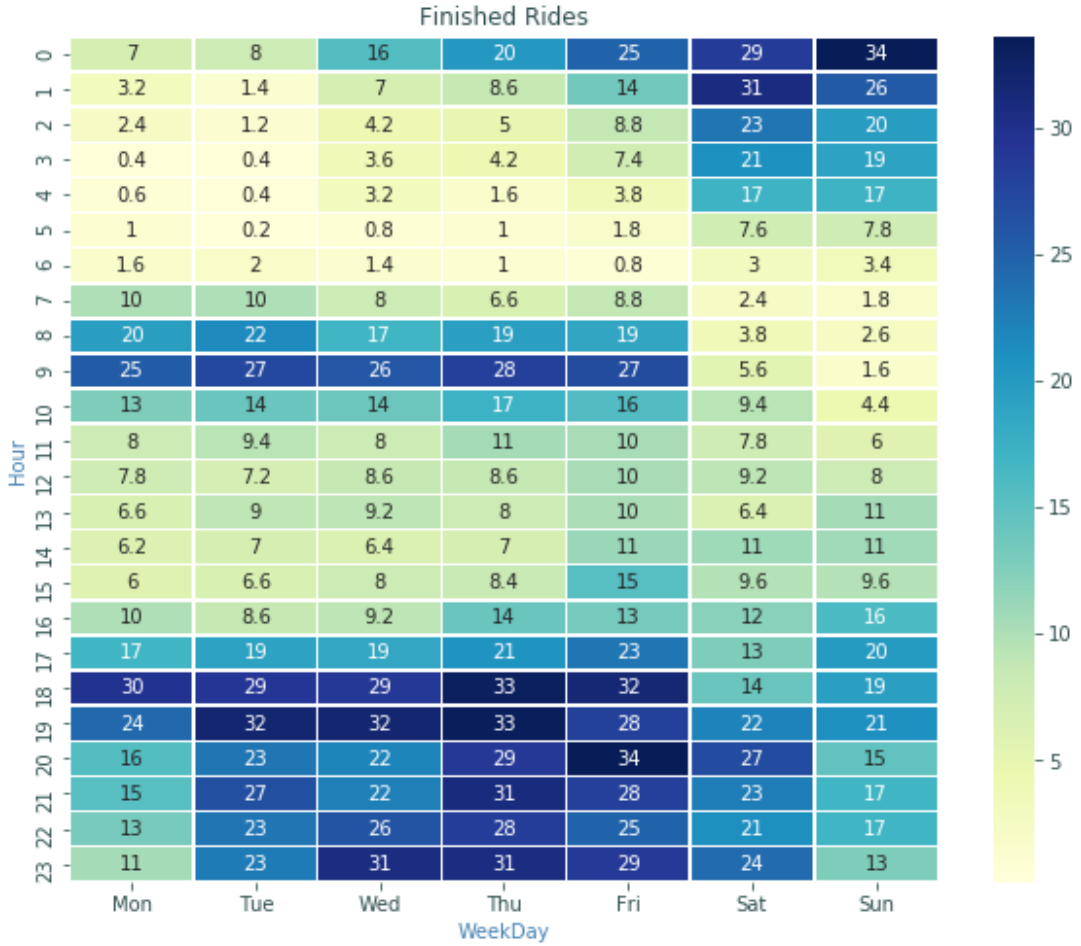
Average ride duration is not provided, but it can be calculated as 'Has booking'/'Finished rides' this info can help to estimate what number of customers can one driver serve during an hour. Average time to get a booking = 18/12= 1.5 hours and average duration of the ride is 6/12=0.5 hours.

Results

Data preparation and Preprocessing

- Two data sources were joined (outer) to a single dataframe
- Missing values filled with 0
- Data consistency checks:
    - Negative values
    - Extreme values (Maximum)

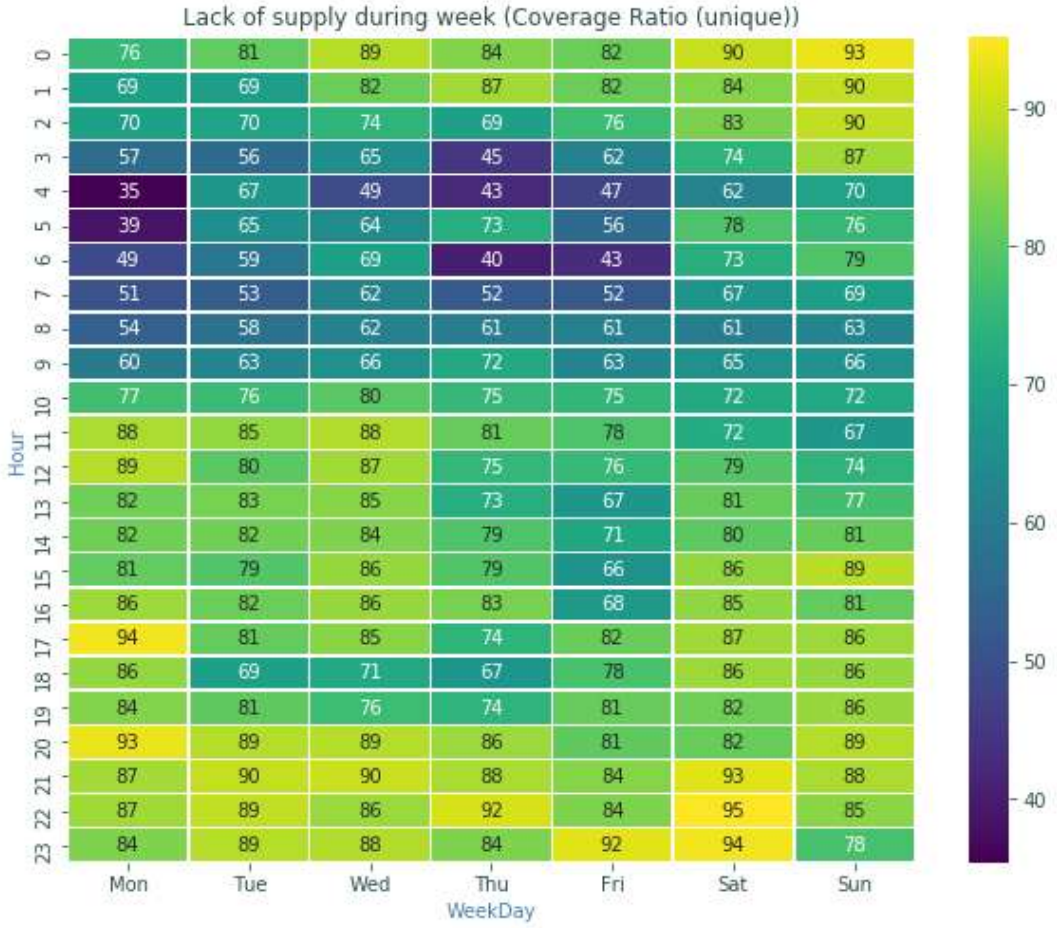
Q1: Show which 36 hours in a week are most undersupplied. Show/describe your decision based on sample data.



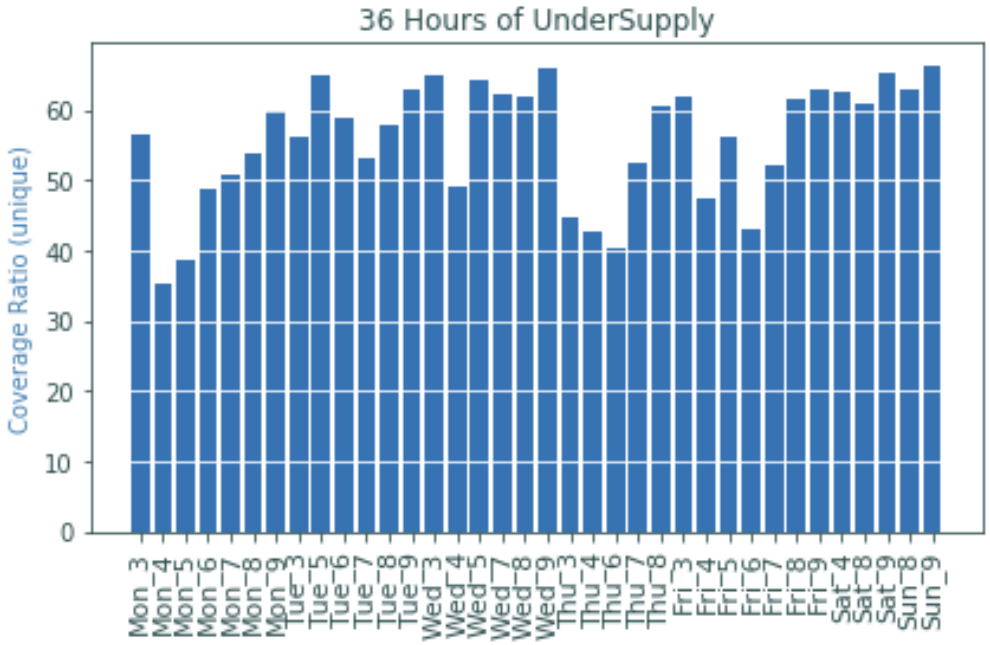
From the data we can observe that most of the rides appear:

- around start and end of "office hours" and at the evenings at working days
- at the evenings and first night hours (coming back from parties) at weekends.

Knowing that most of the drivers prefer to work at the evenings when they have the biggest chance to earn money. That is why the lack of supply appears mainly in the morning hours.

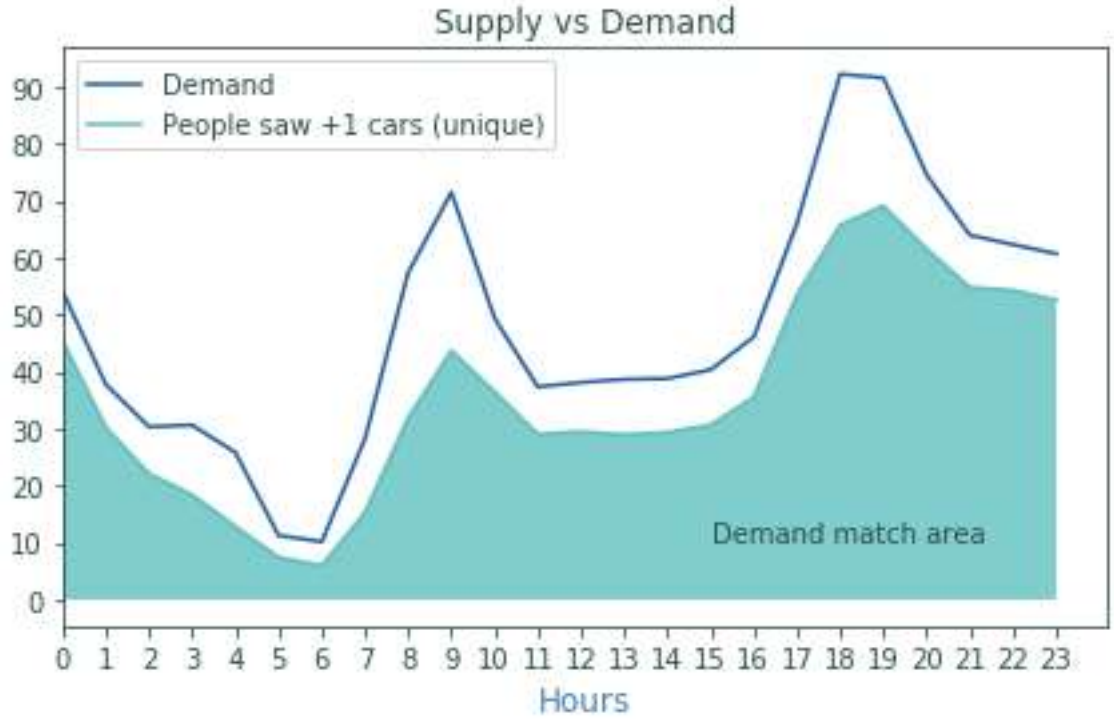


In average there were times no oversupply



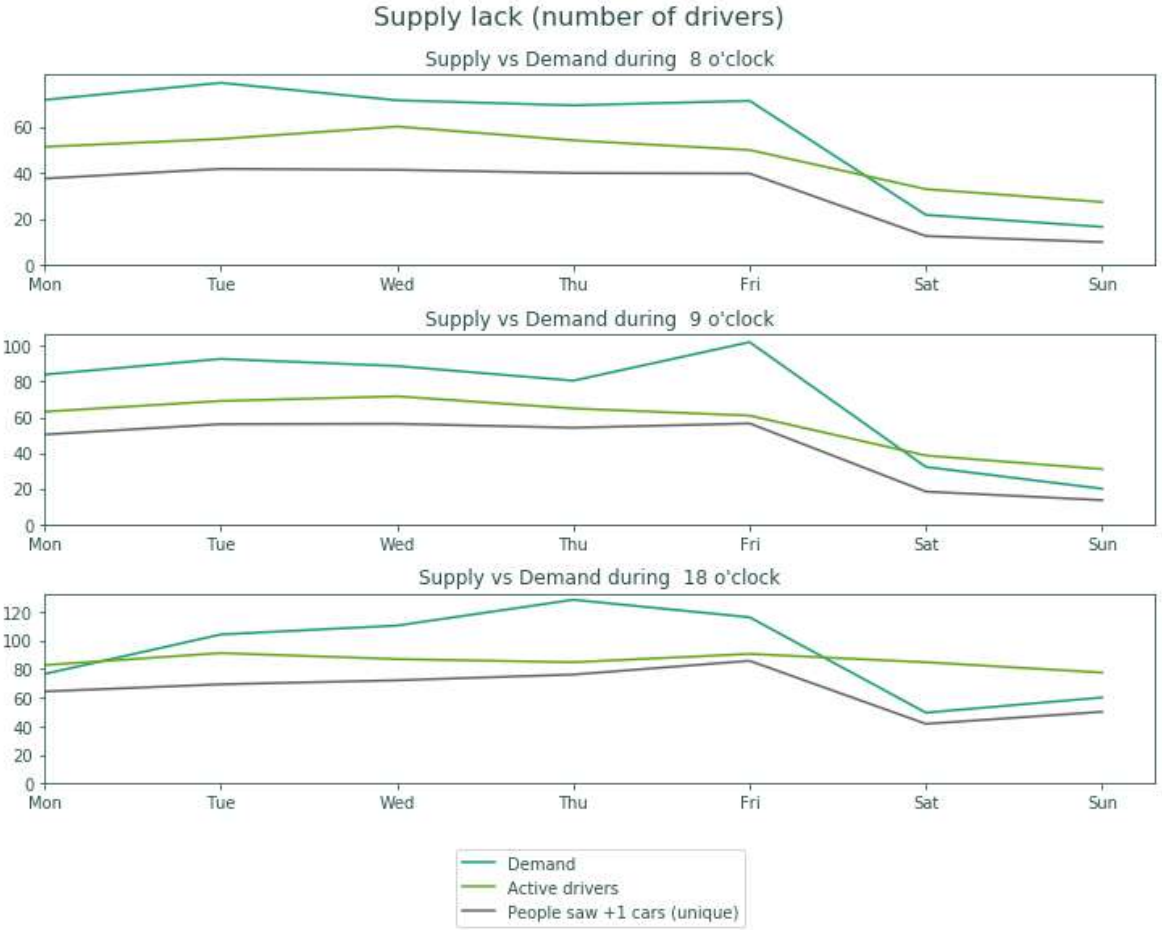
Q2: 24-hour curve of average supply and demand (to illustrate match/mismatch).

The Demand is calculated as a sum of 'People saw 0 cars (unique)' and 'People saw +1 cars (unique)'

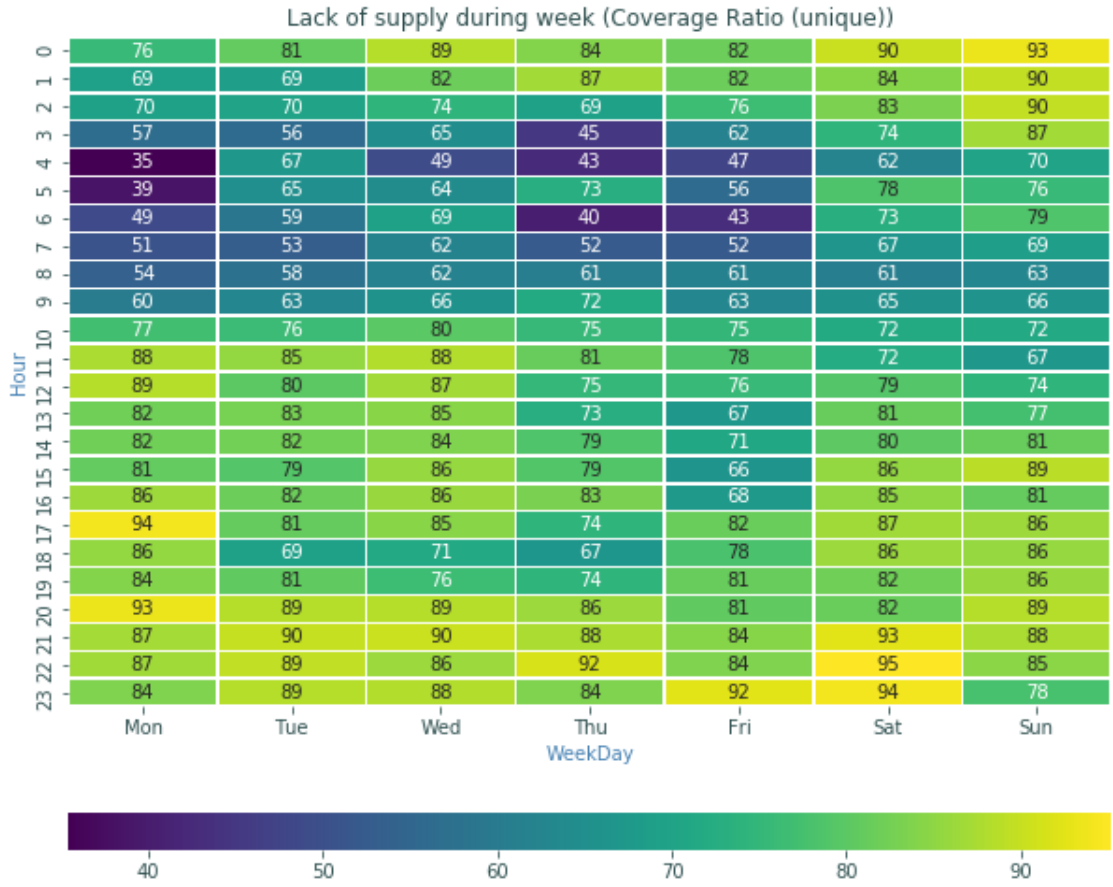


Q3: Visualization of hours where we lack supply during a weekly period. This one we can send to drivers to show when to online for extra hours.

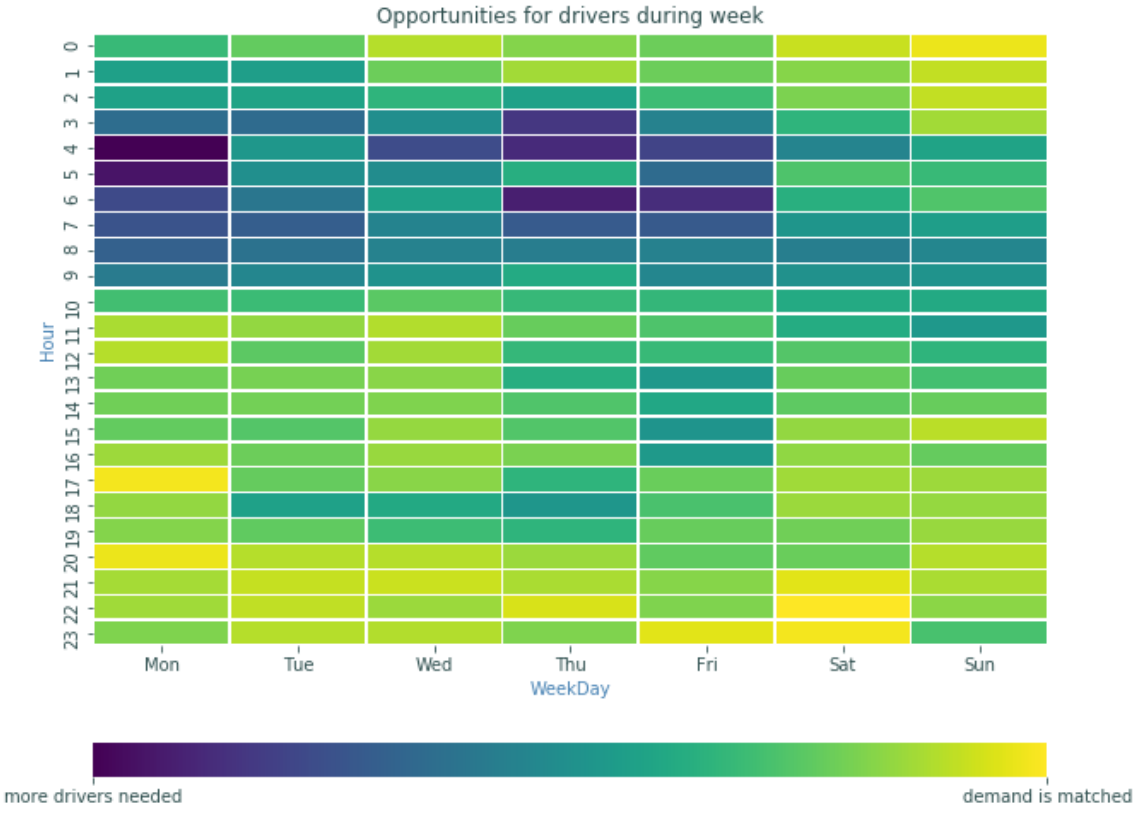
Supply can also be studied based on number of drivers. Not only by extending the number of hours but also increasing the number of active driver during the hours the company can receive the better match.



For the company (internal use)



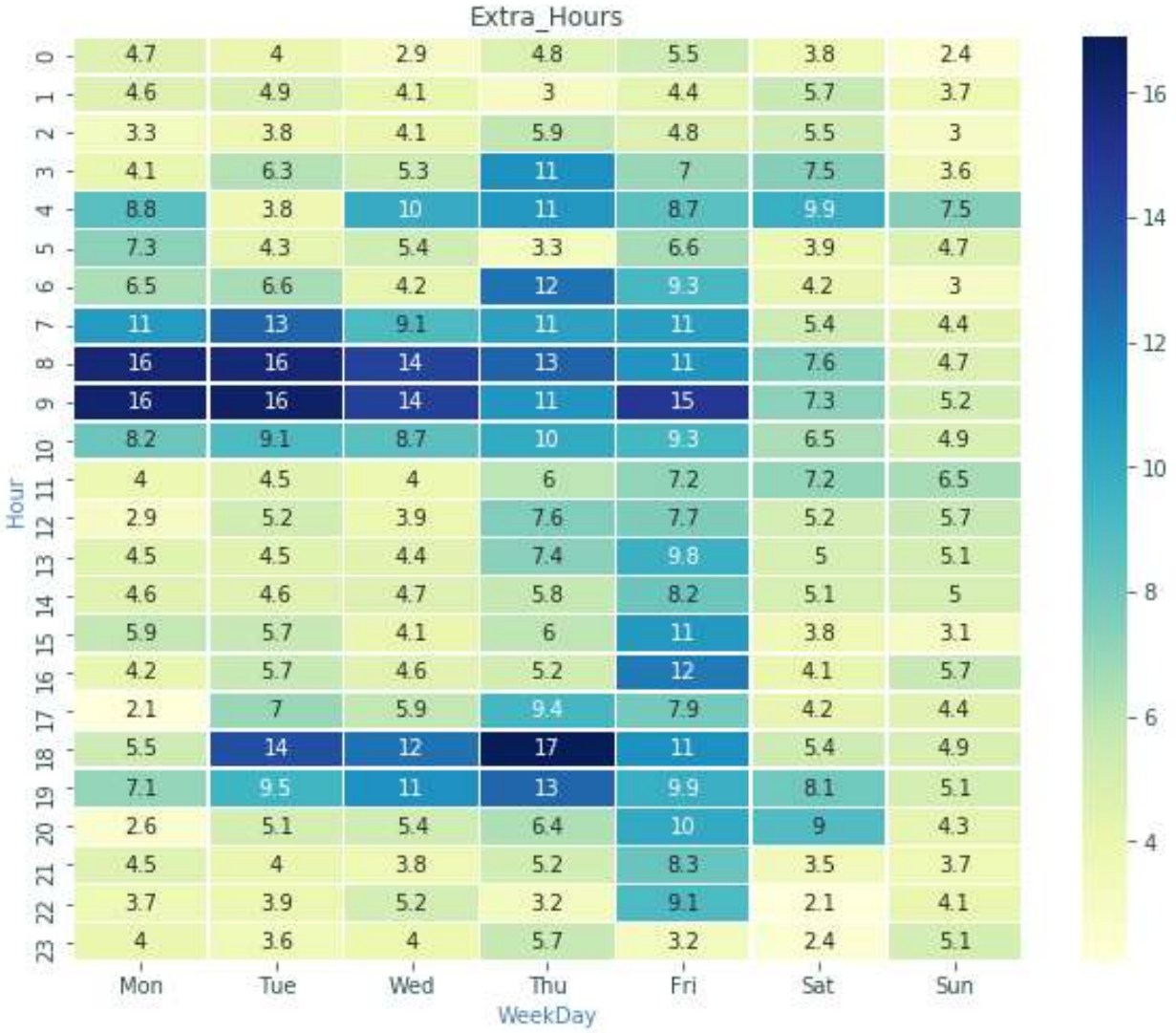
For drivers (external use)



Only the ‘Opportunities for driver during week’ can be shared with drivers so is only minimal amount of data is provided (no ‘data leakage’ to other competitors).

Q4: Estimate number of hours needed to ensure we have a high Coverage Ratio during most peak hours.

$$ExtraHours = Online\ (h) * (\frac{100}{Coverage\ Ratio\ (unique)} - 1)$$



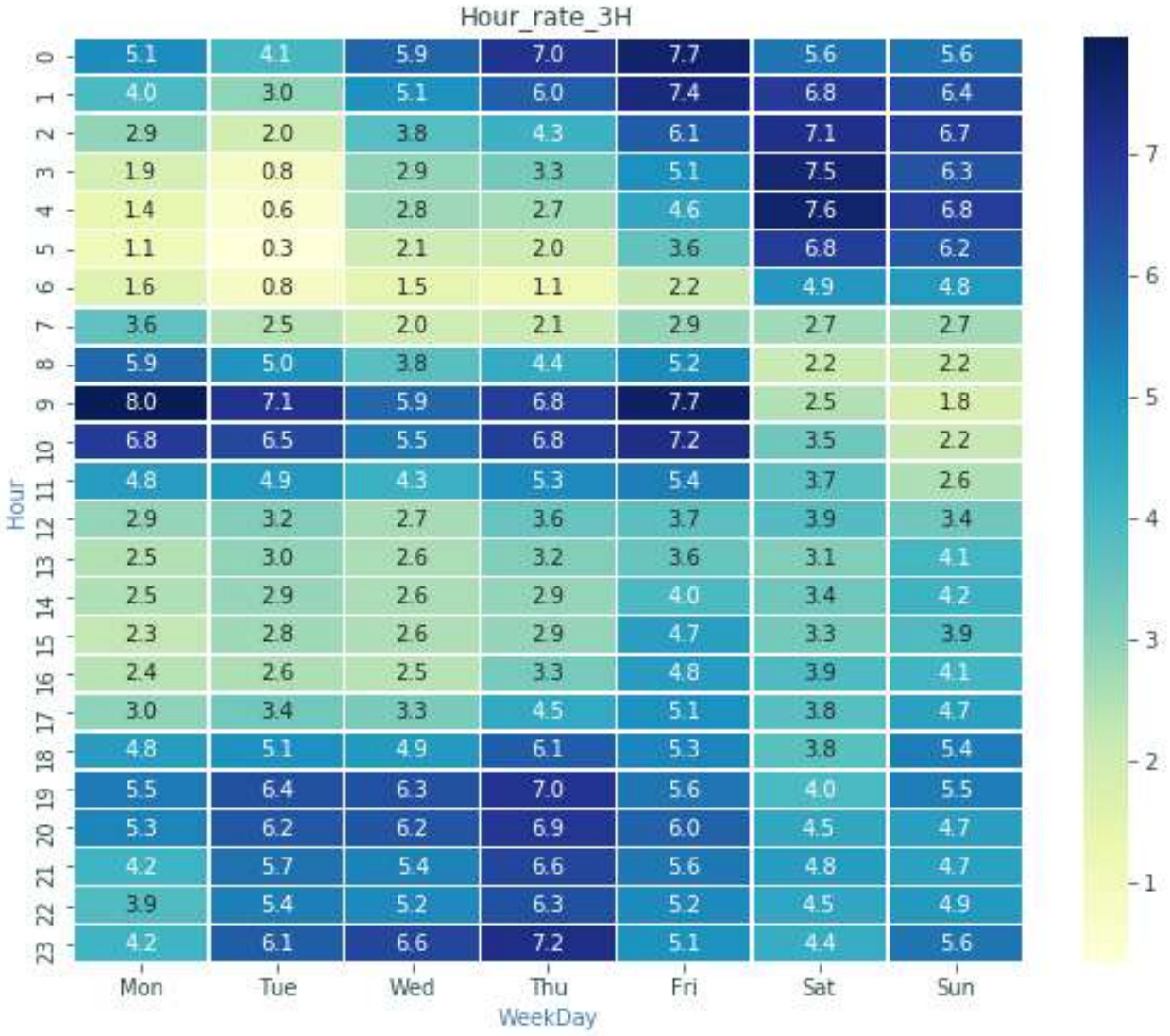
Q5: Calculate levels of guaranteed hourly earnings we can offer to drivers during 36 weekly hours with highest demand without losing money + how much extra hours we want to get to capture missed demand.

a. **Assume that Finished Rides have an average value of €10 (ValuePerRide) (80% (DriverShare) goes to driver, 20% (CompanyShare) is our revenue).**
b. **Assume the same level of demand with increased supply, base it on RPH over 3 hour periods, but with increased supply.**
c. **Assume that with extra hours we will capture "missed coverage" or people attributed to "People saw 0 cars" in demand data.**

Based on the description the hourly rate can be estimated by following formula:

$$Rate\ per\ Hour = \frac{\frac{Online(h) * RPH\_avg\_3H}{CoverageRatio} * 100 * ValuePerRide}{Online(h) + ExtraHours} * DriverShare$$

This Rate includes the DriverShare so if the company is paying the guarantied hourly rate the driver doesn't receive a fee from the ride.



Guaranteed hourly earnings during 36 weekly hours with highest demand

WeekDayName_Hour	WeekDayName	WeekDay	Hour	Extra_Hours	Hour_rate_3H
Mon_8	Mon	0	8	15.97	5.88
Mon_9	Mon	0	9	16.27	7.97
Mon_18	Mon	0	18	5.45	4.8
Mon_19	Mon	0	19	7.09	5.54
Tue_8	Tue	1	8	15.93	5.01
Tue_9	Tue	1	9	16.47	7.07
Tue_18	Tue	1	18	13.84	5.09
Tue_19	Tue	1	19	9.53	6.35
Wed_8	Wed	2	8	14.34	3.77
Wed_9	Wed	2	9	14.45	5.87
Wed_17	Wed	2	17	5.9	3.27
Wed_18	Wed	2	18	12.5	4.89
Wed_19	Wed	2	19	11.31	6.34
Thu_9	Thu	3	9	11.08	6.77
Thu_17	Thu	3	17	9.36	4.49

WeekDayName_Hour	WeekDayName	WeekDay	Hour	Extra_Hours	Hour_rate_3H
Thu_18	Thu	3	18	16.9	6.08
Thu_19	Thu	3	19	12.93	7.02
Thu_20	Thu	3	20	6.37	6.93
Thu_21	Thu	3	21	5.21	6.58
Thu_23	Thu	3	23	5.69	7.21
Fri_8	Fri	4	8	11.06	5.18
Fri_9	Fri	4	9	14.93	7.73
Fri_17	Fri	4	17	7.89	5.09
Fri_18	Fri	4	18	11.34	5.29
Fri_19	Fri	4	19	9.86	5.63
Fri_20	Fri	4	20	10.16	5.97
Fri_21	Fri	4	21	8.28	5.55
Fri_22	Fri	4	22	9.1	5.23
Fri_23	Fri	4	23	3.21	5.07
Sat_0	Sat	5	0	3.76	5.6
Sat_1	Sat	5	1	5.69	6.84
Sat_2	Sat	5	2	5.45	7.15
Sat_3	Sat	5	3	7.5	7.51
Sat_19	Sat	5	19	8.06	3.98
Sat_20	Sat	5	20	9.02	4.53
Sun_0	Sun	6	0	2.4	5.63

Guaranteed hour rate in 36 hours with the highest demand range from 3.27 and 7.97 Euro, average rate is 5.8 Euro.