

Appendix: Adjustments to the Gridwise Data



Appendix to the report

Gig Passenger and Delivery Driver Pay in Five Metro Areas

by Ken Jacobs, Michael Reich, Tynan Challenor, and Aida Farmand • May 2024

<https://laborcenter.berkeley.edu/gig-passenger-and-delivery-driver-pay-in-five-metro-areas/>

The Gridwise trip data provide us with the gross earnings of each driver during P23 time over a two-week period, as well as driver shift times. To estimate each of our earnings measures, we used data from an industry-sponsored study. We also had to account for missing data for some trips and for trips in the data with starting and ending times that did not align with shift times. Moreover, DoorDash trip records were aggregated at the shift level.

To obtain estimates of driver gross earnings during their entire shifts requires also knowing their P1 times during their shifts. We obtain P1 times for each shift by subtracting aggregate P2 and P3 time over the shift from aggregate shift time. To estimate net earnings over an entire shift, we subtract business expenses from gross earnings. To estimate business expenses, we use the IRS expense per mile reimbursement rate and estimate P1, P2, and P3 miles, which are not provided by Gridwise. To obtain these miles, we adjusted P1, P2, and P3 times by average gig passenger driver miles per hour, which we obtained from the industry-sponsored study.

In this Appendix we discuss how we arrived at a cleaned dataset by making these adjustments to the Gridwise data. In Section 1 we describe the Gridwise data. In Section 2 we describe our adjustments for missing data and in Section 3 we discuss our adjustment for trips that do not align with shifts. Making these adjustments reduces our dataset somewhat. In Section 4, we use the results of Sections 2 and 3 and present our adjusted estimates of P1, P2, and P3 times, miles, and miles per hour. In Section 5, we use the results in Section 4 to further adjust our original dataset. Section 6 presents the estimates that result from all these adjustments and compares them to the estimates in our smaller dataset used to estimate parameters for adjustments. As expected, P1 time and miles make up a smaller share of the total in the dataset that includes overlapping trips, given there is no waiting time or miles between such trips.

1. The data

Gridwise provided us with three files: a trip records file, a shift records file, and a monetary incentive file. Each tracked driver activity for the two-week period from January 10, 2022, to January 23, 2022. The trip records file included 63,408 trips by 1,254 unique drivers. The shift records included 10,347 work shifts by 1,223 unique drivers. The incentive file included 4,304 monetary performance incentive records for 657 unique drivers.

These files encompass seven platforms: five delivery network companies—DoorDash, Grubhub, Instacart, Shipt, and Uber Eats—and two passenger network companies—Lyft and Uber. The number of trips for each service: DoorDash (12,004), Grubhub (6,681), Instacart (3,305), Lyft (7,969), Shipt (814), Uber (15,077), and Uber Eats (17,558).

The data include trips in five metros: the nine-county San Francisco Bay Area (6,806); the six-county Greater Los Angeles Region (25,245); King County, Wash., which includes Seattle (6,164); Cook County, Ill., which includes Chicago (14,156); and Suffolk County, Mass., which includes Boston (6,277). The data also included 4,404 trips outside these five metros.

The trip file contains the request time, start time, end time, distance, driver base earnings, driver tip earnings, and driver total earnings for each trip, as well as a unique driver ID. It does not include the time between the end of a trip and the next accepted request. Earnings are recorded at the trip level. The shift file contains the start times, end times, and distances for each shift for each unique driver ID. The companies pay any lump sum incentive earnings and allocate them among all a driver's trips over a two-week period.

2. Missing data

Our earnings estimates depend on having reliable data on time and miles for P1, P2, and P3. As we have mentioned, the trip records include time but not miles for P2 and P3. Moreover, some of the trip files had missing or incomplete data. DoorDash did not report individual trips; instead, it reported only data aggregated at the shift level. As a result, it is not possible to allocate P1, P2, and P3 time in those records. In addition, DoorDash trip records are also missing the distance traveled in each shift.

Trip durations were also missing in some of the Grubhub, Lyft, Shipt, Uber, and Uber Eats records. To estimate these durations, we divided the trip distance by the average P3 miles per hour for each service, in each metro area. This method is imperfect, as it neglects differences in P3 miles per hour among drivers as well as differences in miles per hour during P1 and P2.

Trip distances, but not durations, were also missing in some of the Grubhub, Shipt, and Uber records. In these cases, we estimated the trip distance by multiplying the trip duration by the average P3 miles per hour for each service and metro area. All Shipt records were missing both trip distance and trip duration, making it impossible to impute the missing values. We therefore removed the 40 Shipt drivers from the dataset. The Grubhub and Instacart records did not include P2 times, but Uber Eats did. We used the ratio of P2 to P3 time for Uber Eats by metro to estimate P2 times for Grubhub and Instacart.

3. Misalignment of trips and shifts

In this section, we first discuss how we merged the trip and shift data by driver and excluded trips that began before, or ended after, that driver's shift. After we completed these steps, the data included 54,624 trips, 9,267 shifts, and 1,088 drivers.

The trip and shift datasets do not perfectly align because drivers do not always remember to turn on the Gridwise app before starting their first trip or to close the app after finishing their last trip. There were 11,781 trips that did not fit within the time-window of their shifts. These trips fall into seven categories:

1. Trips that start before and end after a shift, i.e., they engulf the shift
2. Trips that start after all shifts for that driver end
3. Trips that end before all shifts for that driver start
4. Trips in which the end of the trip overlaps the beginning of a shift
5. Trips in which the beginning of the trip overlaps the end of a shift
6. Trips overlapping the beginning and end of two shifts, i.e., straddling two shifts
7. Trips that fall between shifts for that driver and do not overlap any of them

3.1 Truncation of some shifts

Some shifts in the dataset contain gaps of up to 24 hours at the end of the shift and/or long gaps at the beginning or middle of a shift. These gaps likely occur because drivers neglected to log out of the Gridwise, even though they have stopped working. To correct these spurious work gaps, we removed gaps of more than one hour from a shift. This truncation affected 2,564 shifts and 16,118 trips.

After removing shifts in which a trip partly overlaps the beginning or the end of the shift, trips that do not fall into any shifts, and shifts in need of truncation, the dataset includes 24,511 trips, 6,067 shifts, and 935 drivers.

3.2 Overlapping trips

Using those 24,511 trips, the trip file indicates that about half of all trips overlapped in time with another trip. To estimate total trip time in a shift with overlapping trips, we used a recursive algorithm shared by Calum You, a data scientist with Uber. This

algorithm collapses overlapping trip intervals into a single record, allowing us to aggregate total trip time in a shift.<?>

We first calculated the combined P2 + P3 non-overlapping time and then calculated the P2 non-overlapping time. The difference between the two yields the total P3 non-overlapping time.

Estimating overlapping miles is more difficult because drivers do not travel at a uniform speed. Therefore, drivers traverse an unknown number of miles as part of both overlapping trips. We discuss the data cleaning steps for estimating overlapping miles in section 5.

3.3 Calculating P1 share of time and miles

To calculate average P1 time by service (delivery or passenger), we exclude shifts with a trip that is only partly contained in the shift, shifts that need truncation, and shifts with DoorDash trips. The resulting dataset contains 14,604 trips, 2,439 shifts, and 485 drivers. We further exclude shifts with P1, P2, or P3 speeds that exceed 80 miles per hour and shifts with negative P1 distances. We know P1 time and miles for these trips because the shifts and trips perfectly match. Shifts with excessive mile per hour and negative P1 miles reflect discrepancies between miles captured on the Gridwise app and those captured by the service app. Minor discrepancies become important when the start and end of the shift do not align perfectly with the start and end of the trips within the shift.<?>

Table 1. Summary statistics used to estimate P1 time (means)

Trips per shift	Trips per hour	Trip duration (minutes)	Task duration (minutes)	Trip distance (miles)	Shift duration (hours)	Shift distance (miles)
4.2	1.66	23.46	18.46	4.87	2.72	35.88
8.3	2.02	16.2	16.07	7.66	4.08	88.71

<?> See the discussion on StackOverflow. <https://stackoverflow.com/questions/64653134/combine-overlapping-intervals-with-lubridate>

<?> After removing shifts meeting the above criteria, the dataset has 1,669 trips, 626 shifts, and 9 drivers.

Table 2. Summary statistics used to estimate P1 miles (means)

Service	Trips per shift	Trips per hour	Tasks per trip	Tasks per hour	Trip duration (minutes)	Task duration (minutes)	Trip distance (miles)	Shift duration (hours)	Shift distance (miles)
Delivery	2.02	1.51	1.32	1.87	23.48	18.04	4.85	1.62	21.35
Passenger	3.38	1.77	1	1.77	17.1	17.08	8.60	2.14	48.94

3.4 Estimating P2 distance

The Gridwise data contains P2 times, but not P2 distances. To estimate P2 distance, we use a report by the transportation consulting firm Fehr and Peers (2019), which provides vehicle miles traveled for all three trip phases of trips in each of the metros in our dataset, but only for passenger drivers.^{<?>} We use their ratio of P2 time to total time to estimate P2 miles for passenger trips. For delivery trips, we adjusted Fehr and Peers’ reported values by the ratio of delivery to passenger P2 time in each metro. For example, if delivery trips in the San Francisco Bay Area accounted for 80 percent of the P2 time of passenger trips, we adjusted the delivery ratio of P2 miles to total miles by 80 percent. This procedure provides us with overall P2 miles for a shift, but it does not give us P2 miles for individual trips.

4. Shares of time and miles and estimated miles per hour

Here, based on the results of Section 3, we present our estimates of mean P1, P2, and P3 time and mileage shares as well as P1, P2, and P3 miles per hour. In the following tables, the unit of analysis is the shift. For shifts in which a driver completed both passenger and delivery trips, we allocate the shift to the service that accounts for the most miles in the shift.

^{<?>} Fehr & Peers, “Estimated TNC Share of VMT in Six US Metropolitan Regions (Revision 1),” August 6, 2019, https://issuu.com/fehrandpeers/docs/tnc_vmt_findings_memo_08.06.2019.

Table 3. P1, P2, and P3 shares of time

Area	Service	P1	P2	P3	N
California	Delivery	.37	.09	.55	184
California	Passenger	.41	.17	.42	106
California	Total	.38	.12	.50	290
Outside California	Delivery	.37	.14	.49	145
Outside California	Passenger	.38	.18	.44	191
Outside California	Total	.37	.16	.46	336
Total	Delivery	.37	.11	.52	329
Total	Passenger	.39	.17	.43	297
Total	Total	.38	.14	.48	626

Table 4. P1, P2, and P3 shares of miles

Area	Service	P1	P2	P3	N
California	Delivery	.40	.04	.56	184
California	Passenger	.34	.10	.56	106
California	Total	.38	.06	.56	290
Outside California	Delivery	.51	.08	.41	145
Outside California	Passenger	.37	.10	.53	191
Outside California	Total	.43	.09	.48	336
Total	Delivery	.45	.06	.49	329
Total	Passenger	.36	.10	.54	297
Total	Total	.41	.08	.52	626

Table 5. P1, P2, and P3 miles per hour

Area	Service	P1	P2	P3	N
California	Delivery	19.07	13.32	12.45	184
California	Passenger	21.48	16.95	31.03	106
California	Total	19.95	14.65	19.24	290
Outside California	Delivery	25.18	14.98	10.99	145
Outside California	Passenger	22.44	14.81	25.29	191
Outside California	Total	23.62	14.88	19.12	336
Total	Delivery	21.76	14.05	11.81	329
Total	Passenger	22.10	15.57	27.34	297
Total	Total	21.92	14.77	19.18	626

Table 6. P2 + P3 and overall miles per hour

Area	Service	P2+P3	Overall
California	Delivery	11.66	13.05
California	Passenger	26.46	23.86
California	Total	17.07	17.00
Outside California	Delivery	10.55	13.92
Outside California	Passenger	21.85	21.62
Outside California	Total	16.97	18.30
Total	Delivery	11.17	13.43
Total	Passenger	23.50	22.42
Total	Total	17.02	17.70

5. Final adjustments to the data

Here we use the mph results in Section 4 to convert adjusted times to miles and otherwise adjust the data to ensure that shifts match trips in times and miles.

5.1 Trips that overlap shifts

In Section 3 we adjusted the data to eliminate overlapping times. In Section 4 we made corresponding adjustments to eliminate overlapping miles. In this section we use our estimated times, distances, and miles per hour from Section 4 to adjust shifts, times, and miles in our fuller dataset. These are adjustments that we could not make in Section 3. In sections 5.1.1 to 5.1.4, we estimate trip-level P2 distance by multiplying P2 time by average P2 miles per hour for each service type and metro.

5.1.1 Overlap at the beginning of a shift

First, we determine whether only the P3 time of the trip overlaps the beginning of the shift, or whether the P2 + P3 time overlaps. In each case, we determine the percentage of overlapping times. We assume the same percentage of distance overlaps the shift—that is, we assume that the shift distance includes a portion of the overlapping trip. We then extend the shift start time to match the request time of the overlapping trip (i.e., the beginning of P2 time) and add the remaining trip distance to the total shift distance.

5.1.2 Overlap at the end of a shift

We use a similar method to that described in section 5.1.1 to adjust shifts with trips that overlap the end of the shift. After determining whether the P2 time or both P2 and P3 time overlapped, we extend the end time of the shift to match the end time of the overlapping trip. We extended the shift distance to include the portion of the trip that did not overlap the shift.

5.1.3 Trips that start before and end after (engulf) a shift

Trips that engulf shifts have a request time earlier than the shift's start time and an end time later than the shift's end time. We again determined which portion of the trip overlapped which parts of the shift. In most cases this meant determining whether the start time of the trip (i.e., the beginning of P3 time) also preceded the start time of the shift, or whether only the request time did. We determined the percentage overlap

for each segment (P2 and P3 time) and extended the shift start time to match the trip request time and the shift end time to match the trip end time. We then increased the shift distance to include the portion of the trip not captured in the original shift time.

5.1.4 Trips that straddle two shifts

Trips that straddle two shifts overlap the end of one shift and the beginning of the next one. First, we determined the midpoint between the two shifts, which generally fell somewhere in the P3 time of the trip. We then split the trip into two, using the midpoint between the shifts as the boundary. We determined how much of the original trip was allocated to each of the new trips and divided the trip miles and driver earnings accordingly. For example, if the midpoint between the shifts bifurcated the trip into a portion that was 70 percent of the original and another 30 percent, then the miles and earnings were allocated by that ratio. We gave each of the new trips the same proportion of P2 time as the original trip. The situation has now become a trip overlapping the end of a shift and another overlapping the beginning. The shifts were then adjusted using the methods outlined in sections 5.1.1 and 5.1.2.

5.1.5 Adding trips back into the main dataset

After the adjustments described in sections 5.1.1 – 5.1.4, we added the amended trips back into the dataset using the `full_join` function in the **dplyr** package in R. All changes to shift-level information such as the start time, end time, and distance were carried forward into the main dataset.

5.2 Shifts that need truncation

For gaps in shifts of longer than an hour, we substituted the average gap from the dataset accounting for where the gap occurred: at the beginning of a shift, between trips in a shift, or at the end of the shift. These changes reduced the shift time and the shift's P1 time.

Average gaps in shifts that did not need truncation:

- Beginning of shift: 6.83 min
- Between trips: 7.78 min
- End of shift: 13.83 min

We reduced the shift distance and the P1 distance as follows:

- We calculated the new share of P1 time after truncation by dividing the new P1 time by the new shift time.
- We used that value to calculate the factor by which the P1 share of time was reduced—i.e., if P1 time shrank from 50 percent prior to truncation to 40 percent after, then the factor would be 1.25.
- We then reduced the P1 share of miles by the same factor.
- The total shift miles after truncation were then calculated by dividing the P2 and P3 miles (unchanged during truncation) by *(1 - the new P1 share of miles)*.

5.3 Synthetic shifts

The trips that have not yet been accounted for are those that do not overlap an existing shift at all. For these trips, the driver most likely never turned on the Gridwise app during their shift. Using the **dplyr** package in R, drivers were grouped by their driver ID. Any gap between trips of longer than an hour was used to separate shifts.

We determined the total P23 time (accounting for overlapping trips) in the manner of section 3.2. We assigned to each shift the average P1 time of its service type and metro, resulting in a full shift time.

Shift distance was calculated by multiplying the shift time by the average overall miles per hour by service type and metro, leaving only the P1 distance to be determined. The P1 distance for synthetic shifts was the difference of the total shift distance and the P2 + P3 distance. The calculations for non-overlapping P2 + P3 distance are detailed in the next section.

5.4 Overlapping trips

The trip level data reported P3 miles for each trip. It is unknown what proportion of the miles in overlapping trips were traversed in service of both trips. Adding the trip distances together may result in an overestimate of the P3 miles traveled.

For overlapping trips, we multiplied the non-overlapping P2 + P3 time by the average P2 + P3 mph for each service type and metro to estimate non-overlapping P2 + P3 miles. From this value we subtracted the inferred P2 distance for each shift, leaving us with an estimated total P3 shift distance. If this value was less than the sum of overlapping P3 trip distances, then we used it as the P3 shift distance.

We then calculated the P1 shift distance as the difference between the total shift distance and the sum of the P2 and P3 shift distances.

6. Cleaned dataset

6.1 Overview

The resulting dataset after all the above adjustments contains 52,370 trips, 12,850 shifts, and 1,088 drivers.

Our earnings calculations exclude DoorDash trips because its records were missing trip-level information. The cleaned dataset, removing drivers with any DoorDash trips, contains 30,100 trips, 5,222 shifts, and 434 drivers. The sample size for each metro is: Boston (50), Chicago (98), Los Angeles Region (202), San Francisco Bay Area (41), and Seattle (43).

6.2 Adjusted shares of times and miles

Table 7. Adjusted P1, P2, and P3 shares of time (means)

Area	Service	P1	P2	P3
California	Delivery	.30	.09	0.61
California	Passenger	.30	.21	0.49
California	Total	.30	.14	.56
Outside California	Delivery	.28	.18	.54
Outside California	Passenger	.31	.21	.48
Outside California	Total	.30	.20	.51
Total	Delivery	.29	.13	.58
Total	Passenger	.31	.21	.48
Total	Total	.30	.17	.53

Table 8. Adjusted P1, P2, and P3 share of miles (means)

Area	Service	P1	P2	P3
California	Delivery	.47	.05	.48
California	Passenger	.31	.12	.57
California	Total	.40	.08	.52
Outside California	Delivery	.52	.10	.37
Outside California	Passenger	.36	.11	.53
Outside California	Total	.44	.11	.45
Total	Delivery	.49	.07	.43
Total	Passenger	.34	.11	.55
Total	Total	.42	.09	.49

Table 9. Adjusted miles per hour (means)

Area	Service	P1	P23	Overall
California	Delivery	20.03	9.56	12.30
California	Passenger	23.10	23.20	22.77
California	Total	21.37	15.51	16.86
Outside California	Delivery	23.34	9.24	13.11
Outside California	Passenger	25.16	20.03	21.31
Outside California	Total	24.34	15.14	17.60
Total	Delivery	21.38	9.43	12.63
Total	Passenger	24.17	21.56	22.02
Total	Total	22.74	15.34	17.20

6.3 Comparison of unadjusted and adjusted data

In Table 10, we compare for California drivers the unadjusted P1, P2, and P3 times and miles in Tables 3 and 4 for the smaller sample with their adjusted counterparts in the larger sample that we report in Tables 7 and 8. The full adjusted set shows a smaller share of time and miles in P1 compared to the smaller dataset with no overlapping trips, or shifts that do not perfectly match trips. These differences in shares of time and miles are what one would expect, given that where there are overlapping trips, we should see less waiting time and miles not accounted for in trips.

Table 10 Unadjusted and adjusted P1, P2, and P3 times and miles, California drivers

Service	P1 time	P2 time	P3 time	P1 miles	P2 miles	P3 miles
Delivery unadj.	.37	.09	.55	.40	.04	.56
Delivery adj.	.30	.09	.61	.34	.10	.56
Passenger unadj.	.41	.17	.42	.47	.05	.48
Passenger adj.	.30	.21	.49	.31	.12	.57