MOIA CASE STUDY RESULTS

Marcos Daniels

TASK 1: PUNCTUALITY

PUNCTUALITY RATE RATHER LOW, NO MAIN FACTOR FOR DELAYS IDENTIFIABLE, MORE ACCIDENTS TO BE EXPECTED IN FUTURE

SUMMARY

Overall, 50% of shifts had an average time difference of around zero while the punctuality rate was only around 51% for total time difference. More than 70% of all shifts were completed at least 5 mins earlier.

Comparing the areas, shifts in area 94fg938d9h were much more punctual (36% in total). Among the drivers, only a small minority of them (around 11%) have a satisfactory time difference saldo. Almost 14% of all drivers worked at least 5 minutes too little in the period.

Shift starts are most punctual on Fridays while most Monday shifts do not end on time planned. Evening shifts start the most punctual and more than every second night shift is ended on time.

Regression shows that none of the available factors has a significant influence on a driver's performance. However, drivers that have a time saldo above the positive threshold have a slightly lower accident probability.

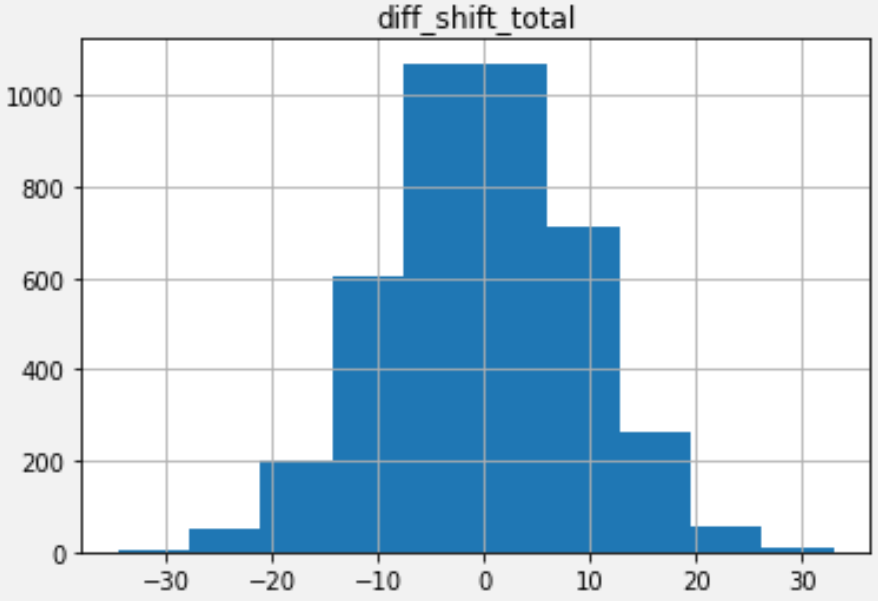
The ARIMA forecast shows that total time difference of all shifts will be negative in the following three days. It also indicates that accident probability will increase within the following three days.

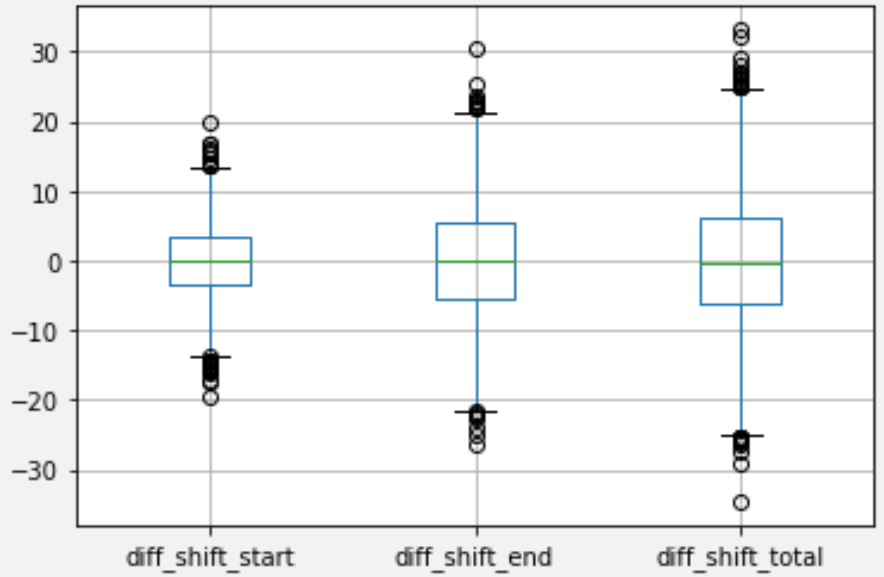
SHIFTS

TIME DIFFERENCE NORMALLY DISTRIBUTED; AVERAGE DIFFERENCES ALMOST AT ZERO

DATA SET LENGTH, DISTRIBUTION & BOX PLOTS

Data on **4,044 shifts** from **three months** (Jan to Mar 2021)





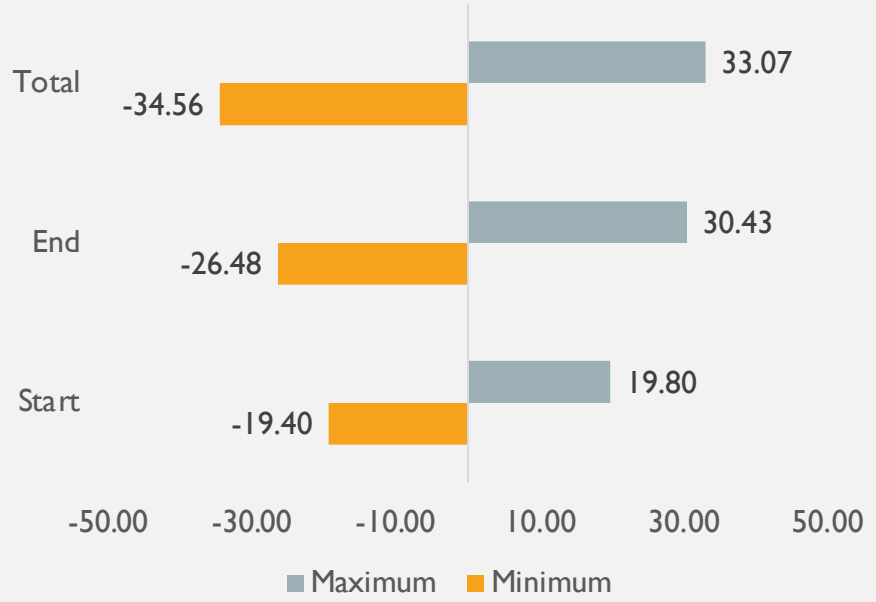
Reading aid: The left chart visualizes the number of total time differences (y-axis) per amount of time (x-axis).

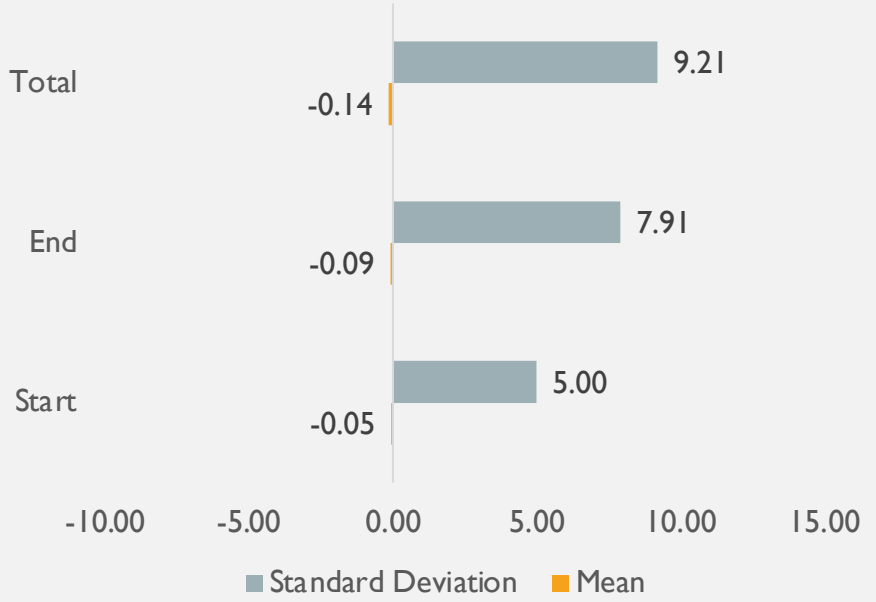
BASE: ALL SHIFTS

Reading aid: The box plots visualize basic measures like median (green line), quartiles (horizontal lines), minimum/maximum (heads) and outliers (bubbles).

TOTAL MAXIMUM DELAY WAS AROUND 33 MINS; AVERAGE DELAY WAS AROUND 9 MINS

MAXIMUMS, MINIMUMS, STANDARD DEVIATIONS & MEANS





BASE: ALL SHIFTS

FOR TOTAL, 50% OF SHIFTS HAD AN AVERAGE TIME DIFFERENCE OF AROUND ZERO; 75% OF SHIFTS HAD AN AVERAGE DIFF BELOW 6.15 MINS

Reading aid: For instance, the first quartile is the value of the 25th percentile. The bottom quarter of the scores fall below this value, while three-quarters fall above it.

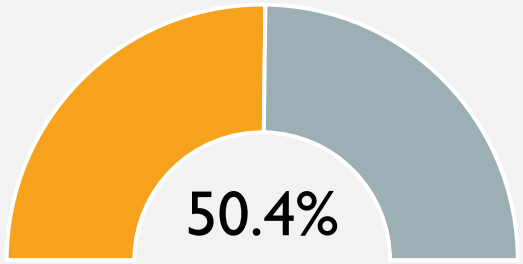
BASE: ALL SHIFTS

DRIVERS WORK LESS THAN PLANNED: -9.58 HOURS DIFFERENCE IN TOTAL; PUNCTUALITY RATE WAS ONLY AROUND 51% IN TOTAL

SALDOS & PUNCTUALITY RATES

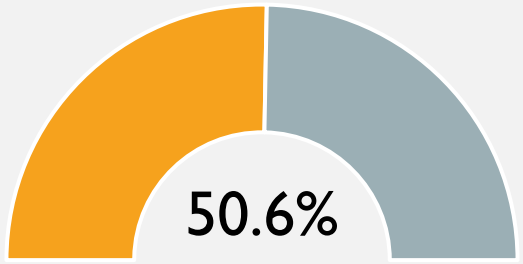
Saldo positive delay:
10.13 days

Punctuality Rate Total



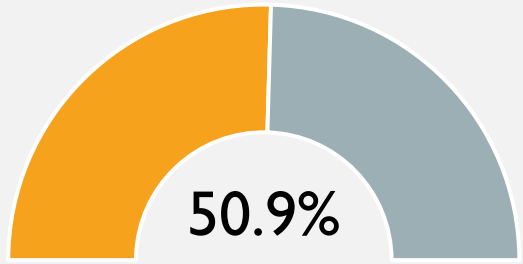
Saldo negative delay:
-10.53 days

Punctuality Rate End



Saldo total:
-9.58 hours

Punctuality Rate Total

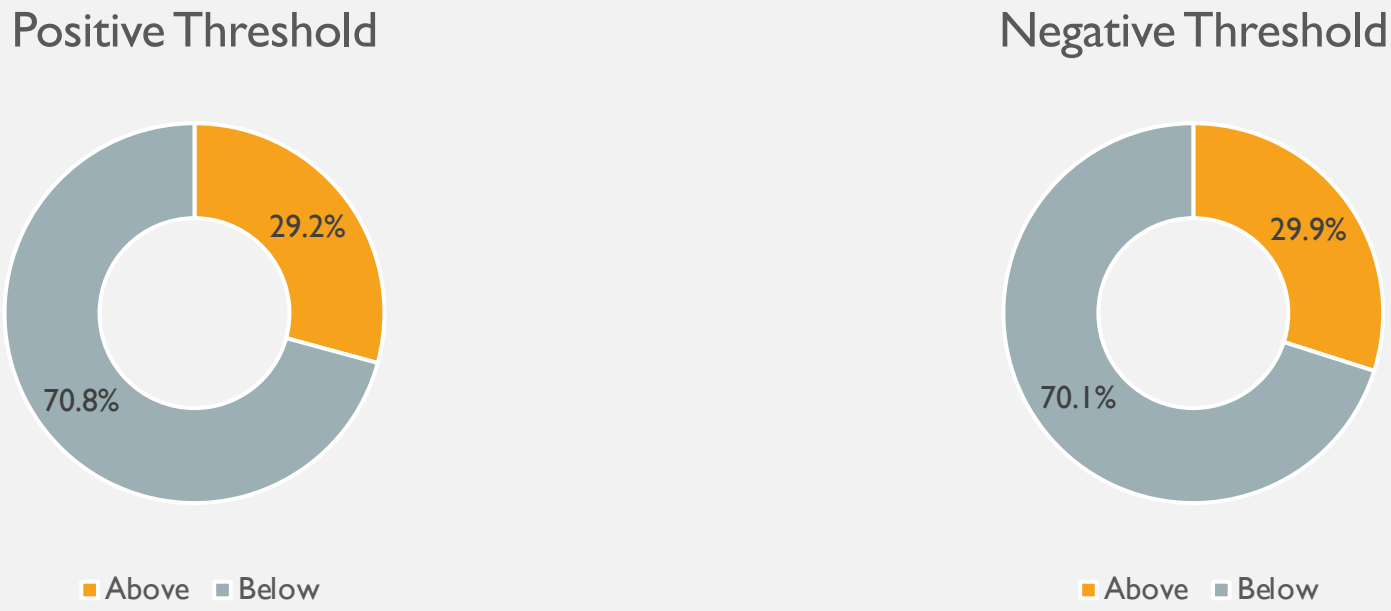


BASE: ALL SHIFTS, TOTAL DIFFERENCE

MORE THAN 70% OF ALL SHIFTS WERE COMPLETED AT LEAST 5 MINS EARLIER

SHARE OF SHIFTS ABOVE/BELOW THRESHOLD

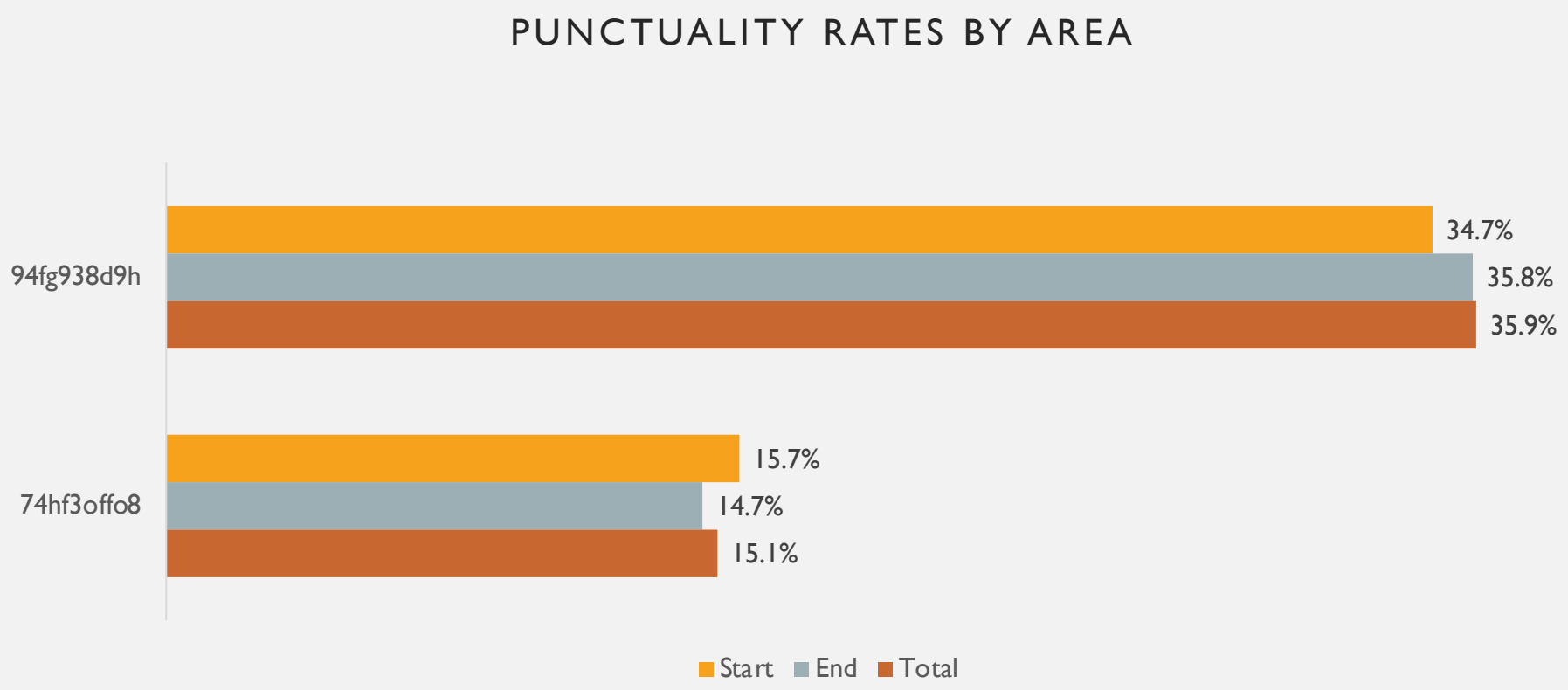
Reading aid: The total time difference of 29.2% of all shifts ranks above the threshold of 5 minutes, that means the completion of those shifts took longer than planned.



BASE: ALL SHIFTS, TOTAL DIFFERENCE; THRESHOLD: 5 MINS

AREAS

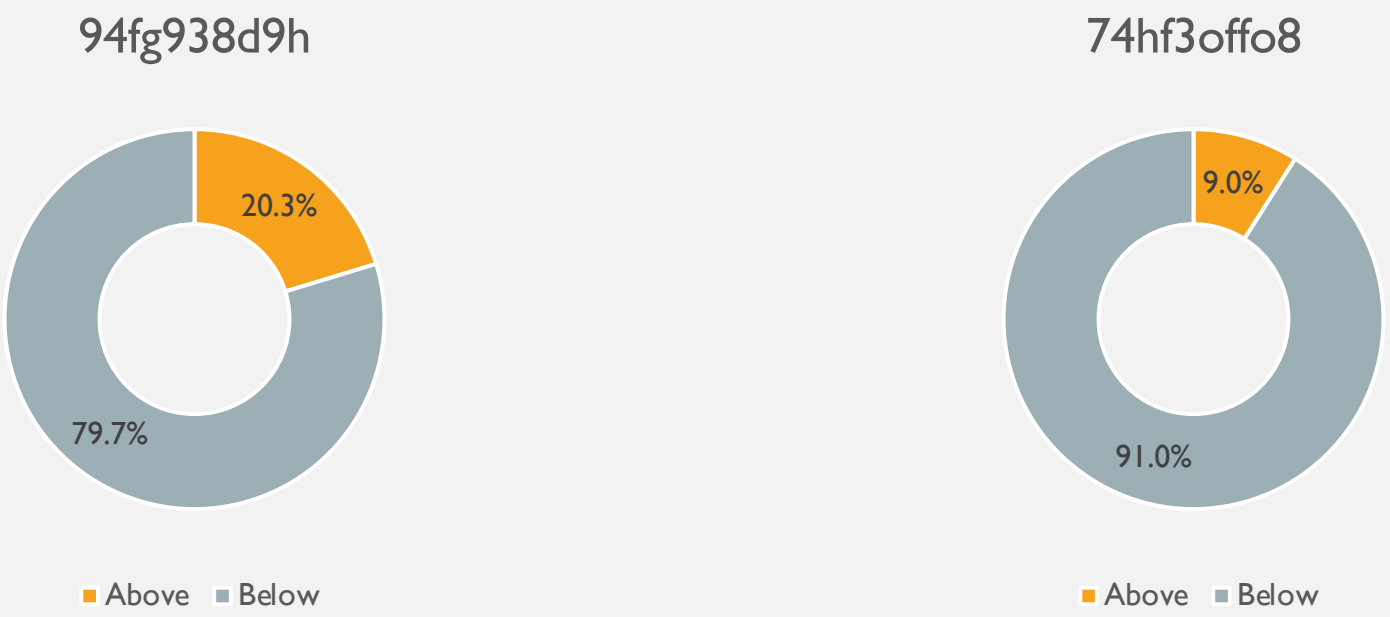
SHIFTS IN AREA WITH ID 94FG938D9H WERE COMPLETED MUCH MORE PUNCTUAL



BASE: ALL SHIFTS, TOTAL DIFFERENCE

MORE THAN 90% OF SHIFTS IN AREA WITH ID 74HF3OFFO8 WERE COMPLETED AT LEAST 5 MINS EARLIER THAN PLANNED

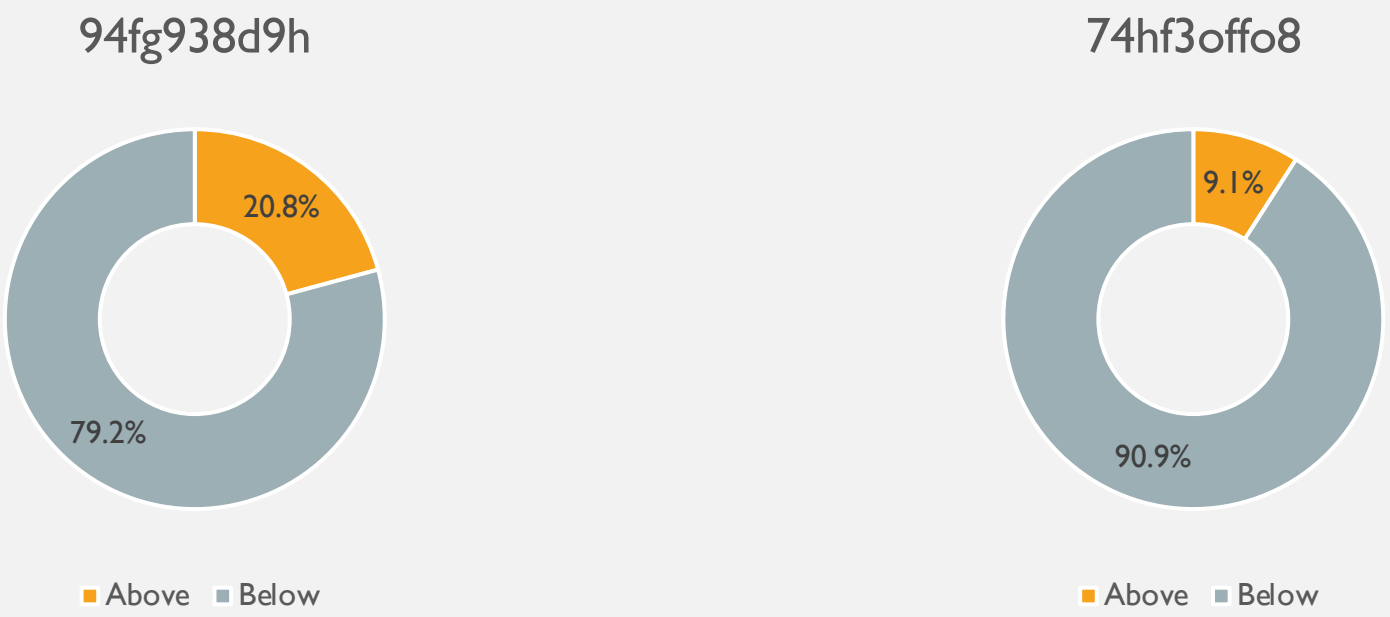
SHARE OF SHIFTS ABOVE/BELOW POSITIVE THRESHOLD BY AREA



BASE: ALL SHIFTS, TOTAL DIFFERENCE; THRESHOLD: 5 MINUTES

EVERY FIFTH SHIFT IN AREA WITH ID 94FG938D9H TOOK MORE THAN 5 MINUTES LATER THAN PLANNED

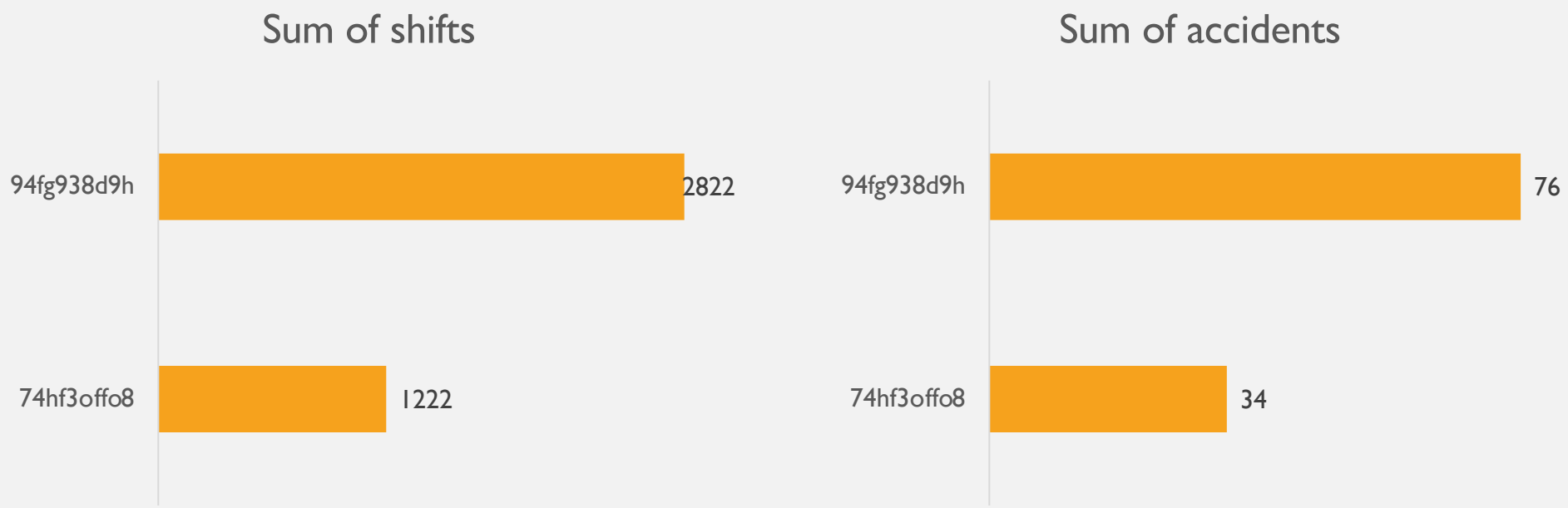
SHARE OF SHIFTS ABOVE/BELOW NEGATIVE THRESHOLD BY AREA



BASE: ALL SHIFTS, TOTAL DIFFERENCE; THRESHOLD: 5 MINS

MOST SHIFTS TOOK PLACE IN 94FG938D9H; MORE THAN TWICE THE AMOUNT OF ACCIDENTS OCCURRED IN THE SAME AREA

SUM OF SHIFTS AND ACCIDENTS BY AREA

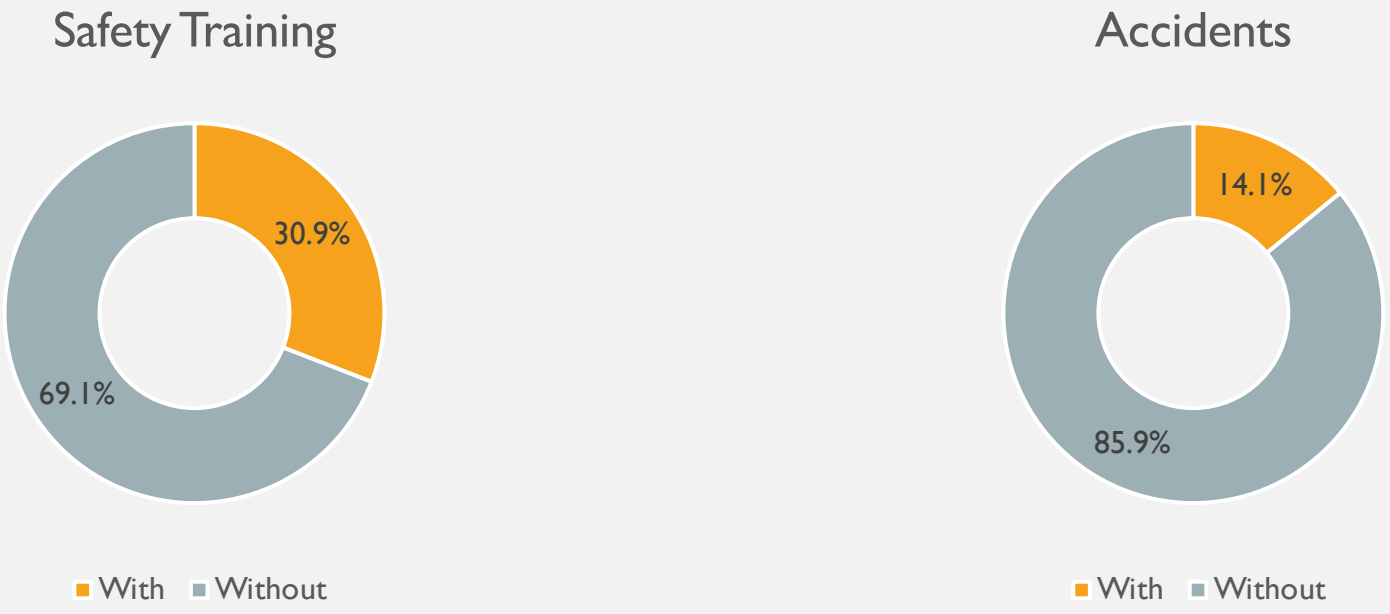


BASE: ALL SHIFTS

DRIVERS

LESS THAN ONE THIRD OF DRIVERS PARTICIPATED IN A TRAINING; 14% HAD AT LEAST ONE ACCIDENT IN THE PERIOD

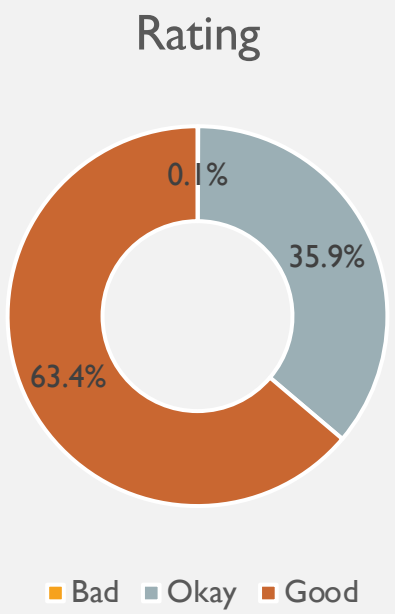
SHARE OF DRIVERS WITH SAFETY TRAINING AND ACCIDENTS

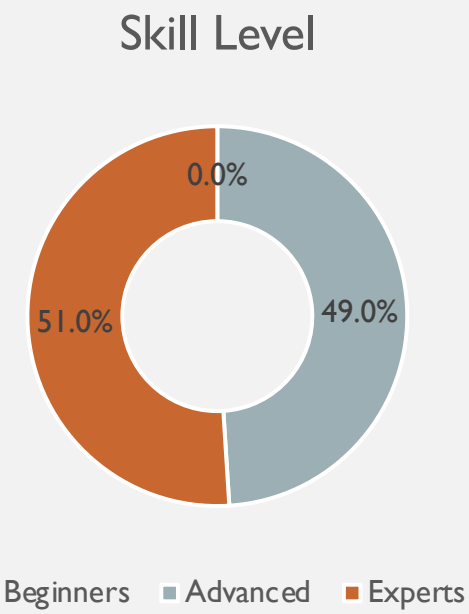


BASE: ALL DRIVERS

SIX OUT OF TEN DRIVERS HAD GOOD RATINGS; HALF OF THEM ARE EXPERTS BASED ON COMPLETED SHIFTS

SHARE OF DRIVERS BY RATING AND SKILL LEVEL

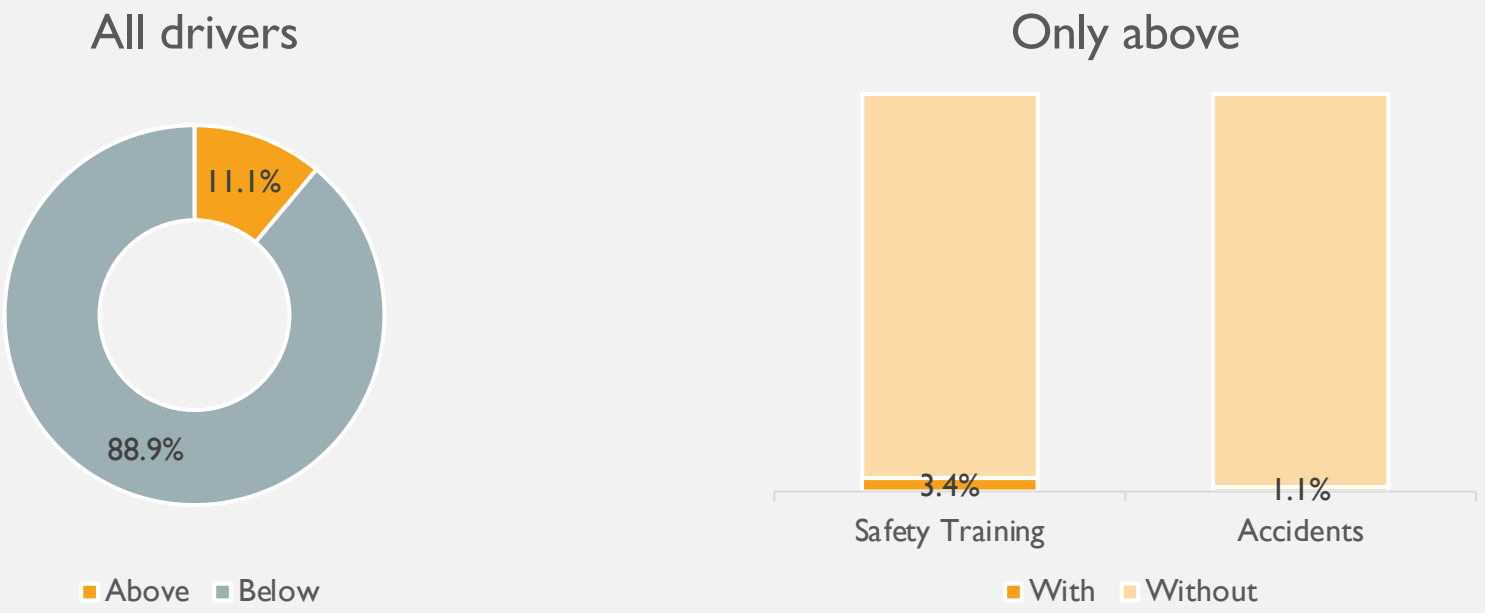




BASE: ALL DRIVERS; S. METHODOLOGY FOR CATEGORY DEFINITIONS

SMALL MINORITY OF DRIVERS HAVE A SATISFACTORY TIME DIFFERENCE SALDO; TRAININGS AND NUMBER OF ACCIDENTS NO FACTOR

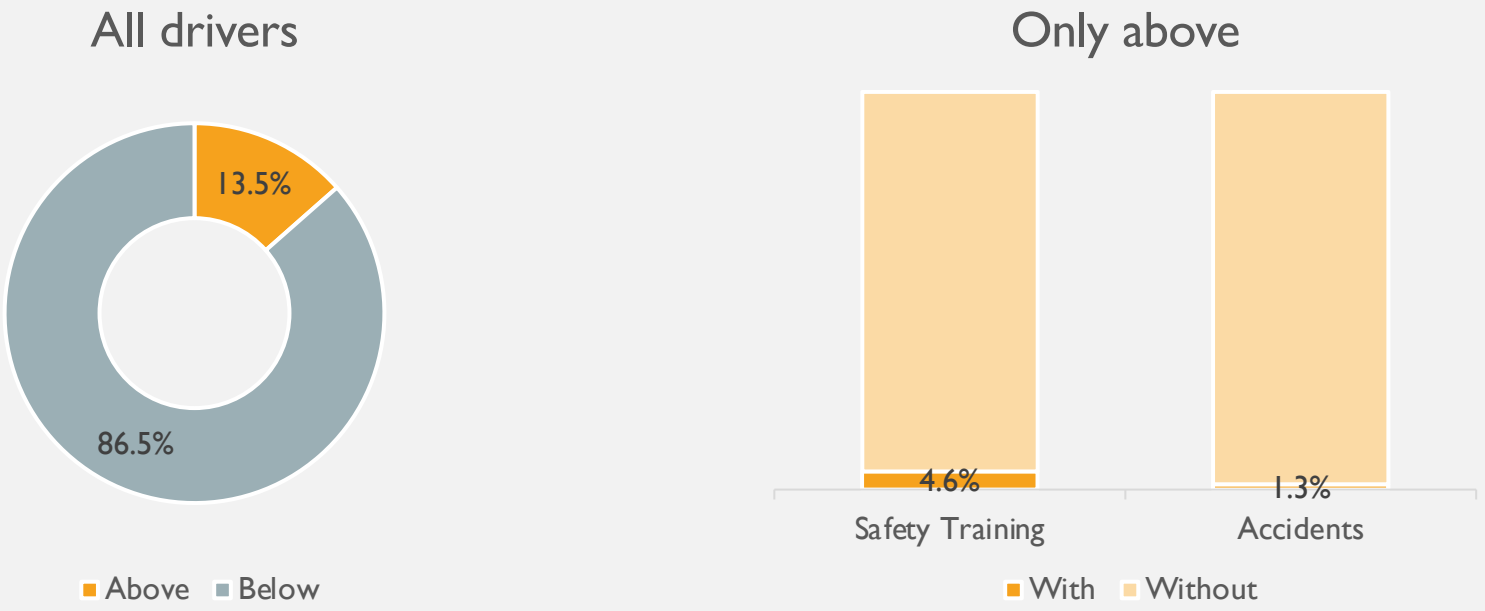
SHARE OF DRIVERS ABOVE/BELOW POSITIVE THRESHOLD & SHARE OF DRIVERS ABOVE POSITIVE THRESHOLD WHO PARTICIPATED IN A TRAINING AND HAD AT LEAST ONE ACCIDENT



BASE: ALL DRIVERS, TOTAL DIFFERENCE (LEFT); DRIVERS WITH TOTAL TIME DIFFERENCE ABOVE POSITIVE THRESHOLD (RIGHT)

ALMOST 14% OF ALL DRIVERS WORKED AT LEAST 5 MINUTES TOO LITTLE IN THE PERIOD

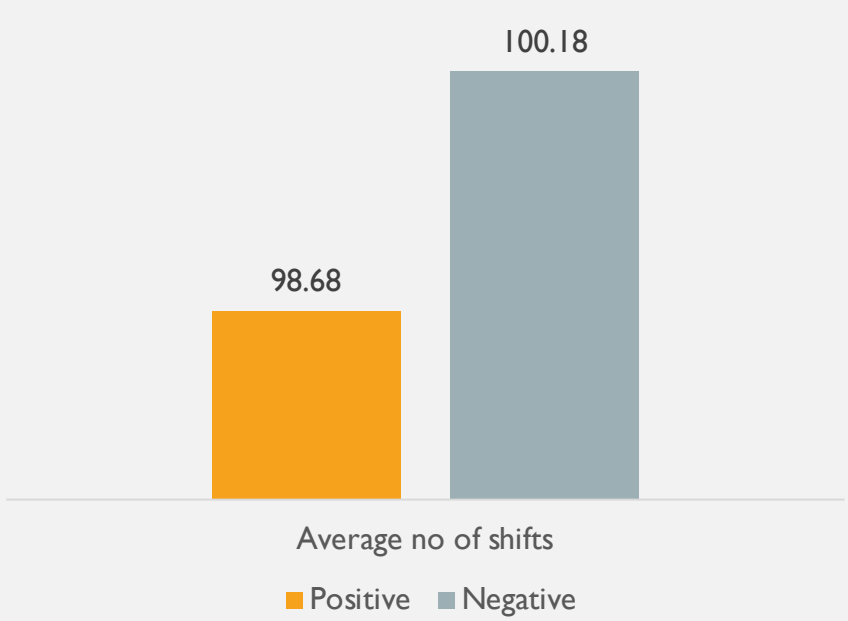
SHARE OF DRIVERS ABOVE/BELOW NEGATIVE THRESHOLD & SHARE OF DRIVERS ABOVE NEGATIVE THRESHOLD WHO PARTICIPATED IN A TRAINING AND HAD AT LEAST ONE ACCIDENT

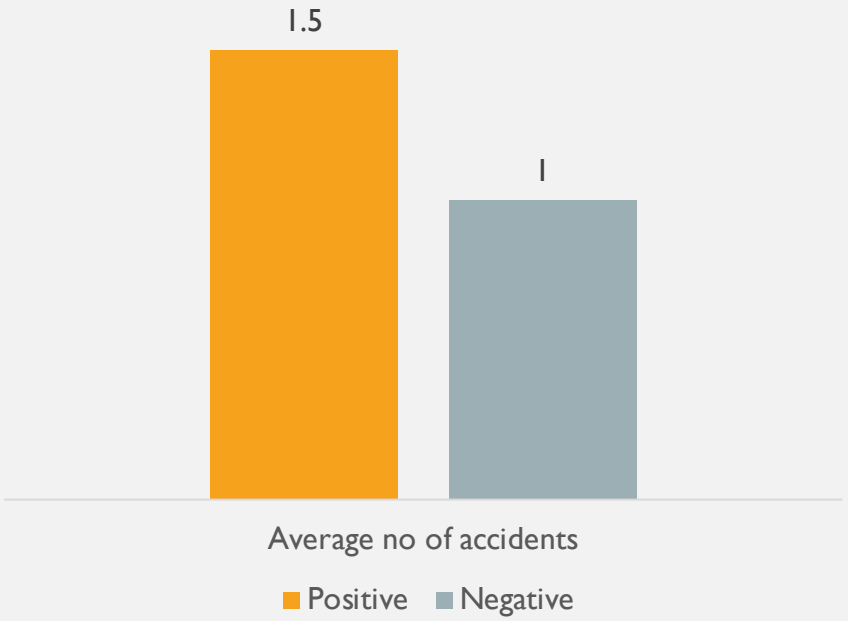


BASE: ALL DRIVERS, TOTAL DIFFERENCE (LEFT); DRIVERS WITH TOTAL TIME DIFFERENCE ABOVE NEGATIVE THRESHOLD (RIGHT)

DRIVERS WHO WORK MORE COMPLETED LESS SHIFTS AND HAD SLIGHTLY MORE ACCIDENTS ON AVERAGE

DRIVERS ABOVE POSITIVE VS NEGATIVE THRESHOLD (1/4) – SHIFTS AND ACCIDENTS

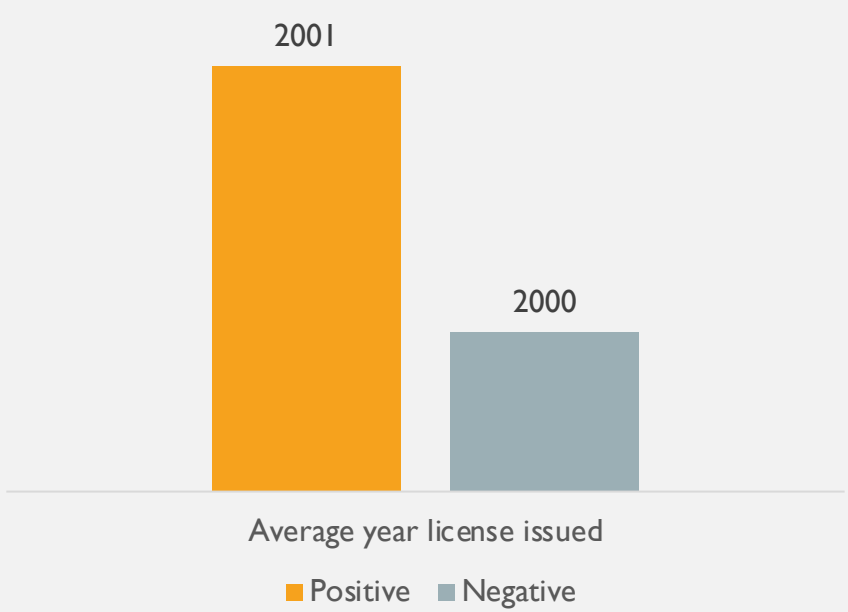


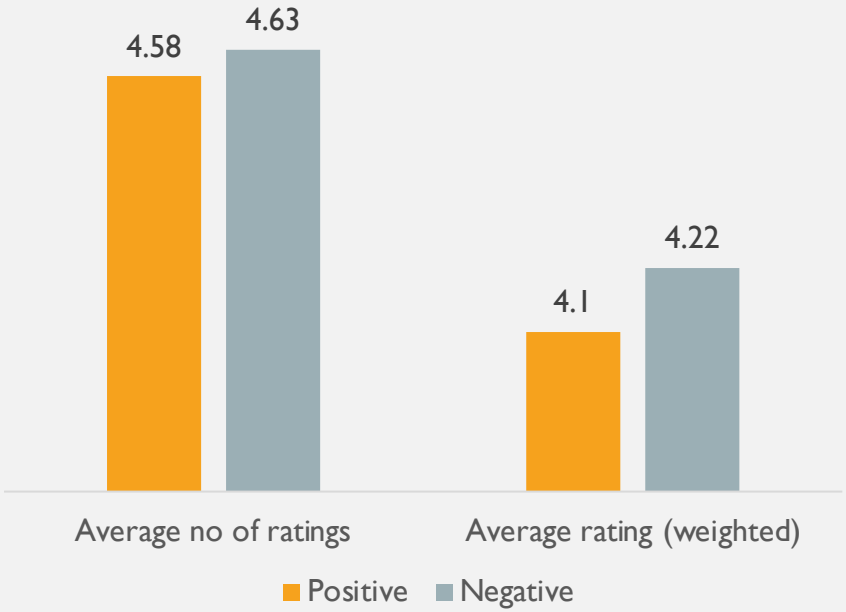


BASE: DRIVERS ABOVE THRESHOLD, TOTAL DIFFERENCE

NO SIGNIFICANT DIFFERENCE BETWEENS DRIVER GROUPS WITH REGARDS TO LICENSE YEAR AND RATINGS

DRIVERS ABOVE POSITIVE VS NEGATIVE THRESHOLD (2/4) – YEAR LICENSE ISSUED & RATINGS

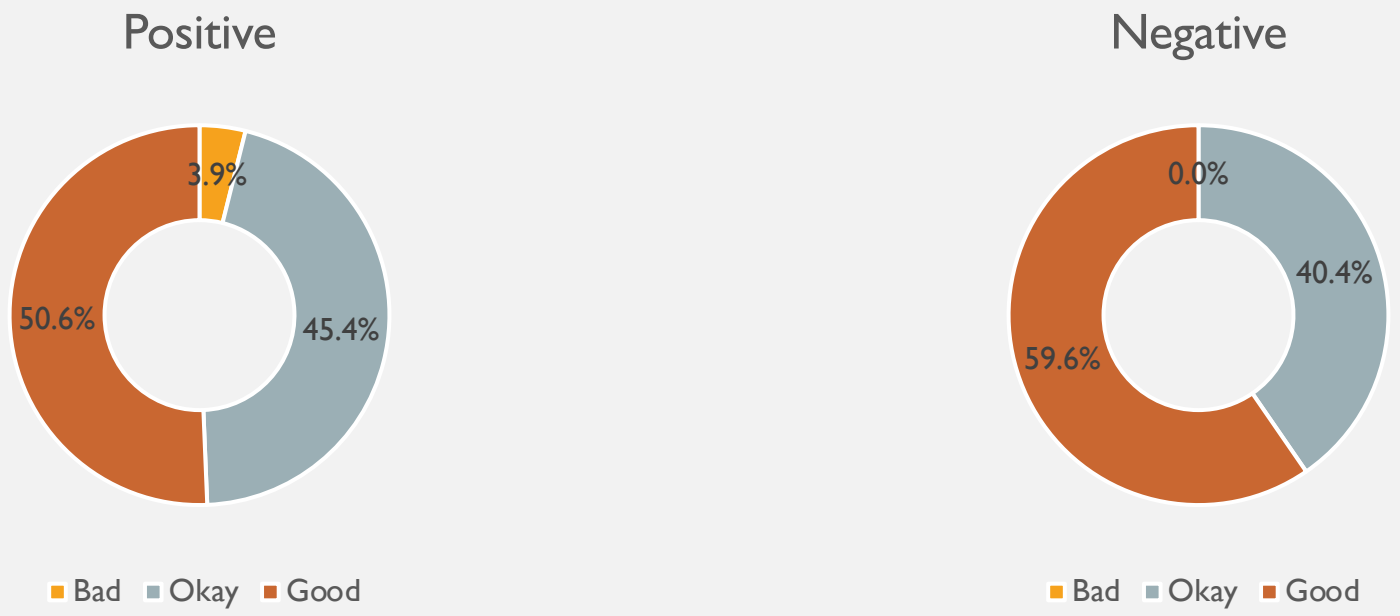




BASE: DRIVERS ABOVE THRESHOLD, TOTAL DIFFERENCE; WEIGHTING OUT OF 7.5 (S. METHODOLOGY)

DRIVERS WHO WORK LESS ARE RATED BETTER THAN DRIVERS THAT WORK MORE

DRIVERS ABOVE POSITIVE VS NEGATIVE THRESHOLD (3/4) – RATING CATEGORY



BASE: DRIVERS ABOVE THRESHOLD, TOTAL DIFFERENCE; S. METHODOLOGY FOR CATEGORY DEFINITIONS

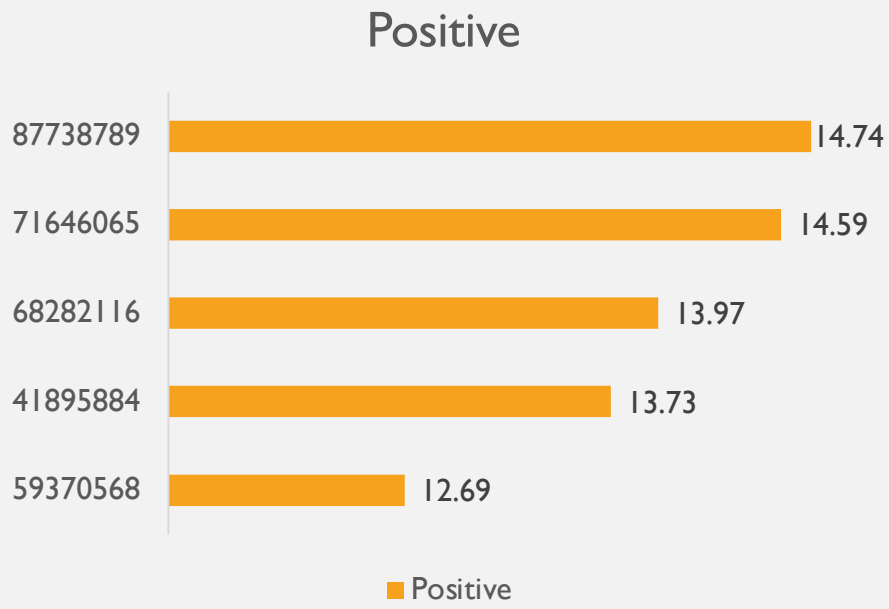
DRIVERS WHO WORK LESS ARE MORE OFTEN EXPERTS BASED ON COMPELTED SHIFTS

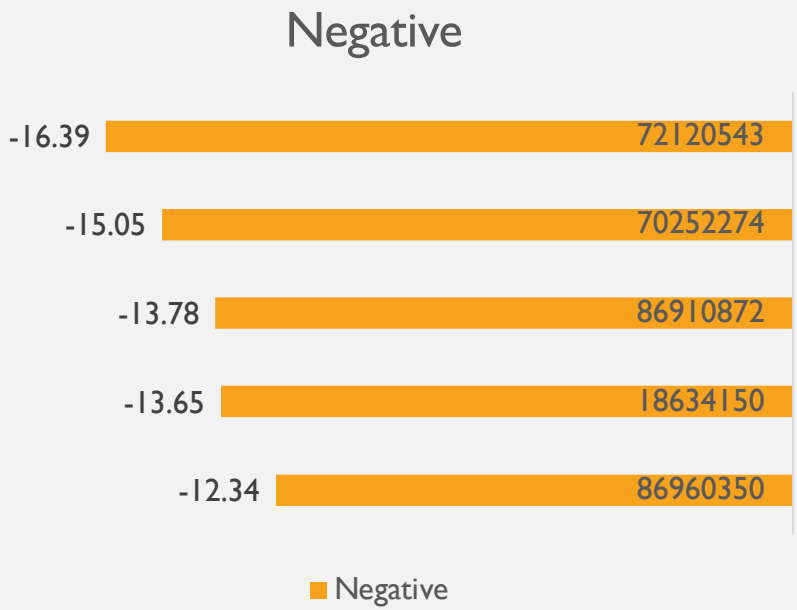
DRIVERS ABOVE POSITIVE VS NEGATIVE THRESHOLD (4/4) – SKILL LEVEL

BASE: DRIVERS ABOVE THRESHOLD, TOTAL DIFFERENCE; S. METHODOLOGY FOR CATEGORY DEFINITIONS

DRIVER WITH ID 87738789 WORKS AROUND 15 MINS MORE ON AVERAGE; DRIVER 72120543 DOES THE OPPOSITE AT AROUND 16 MINS

TOP 5 DRIVERS BY AVERAGE TIME DIFFERENCE

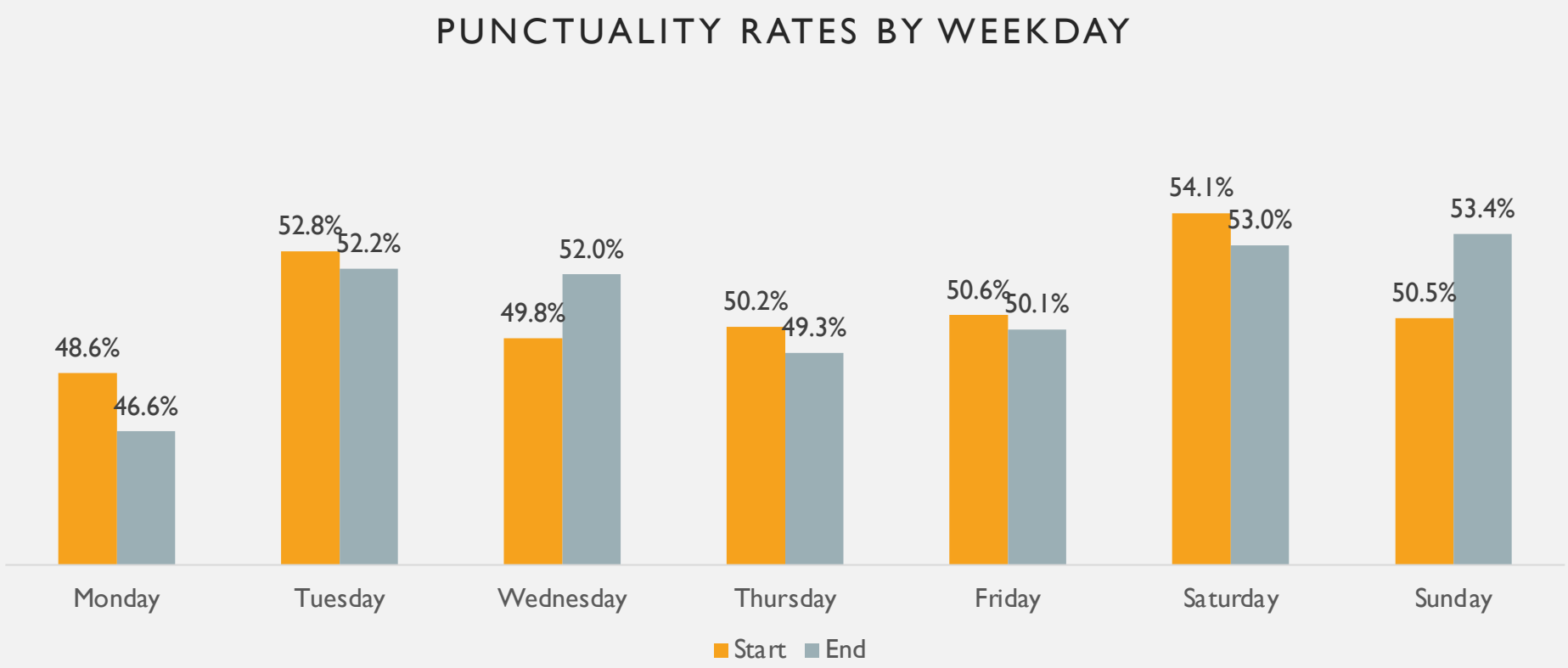




BASE: ALL DRIVERS, TOTAL DIFFERENCE

WEEKDAYS

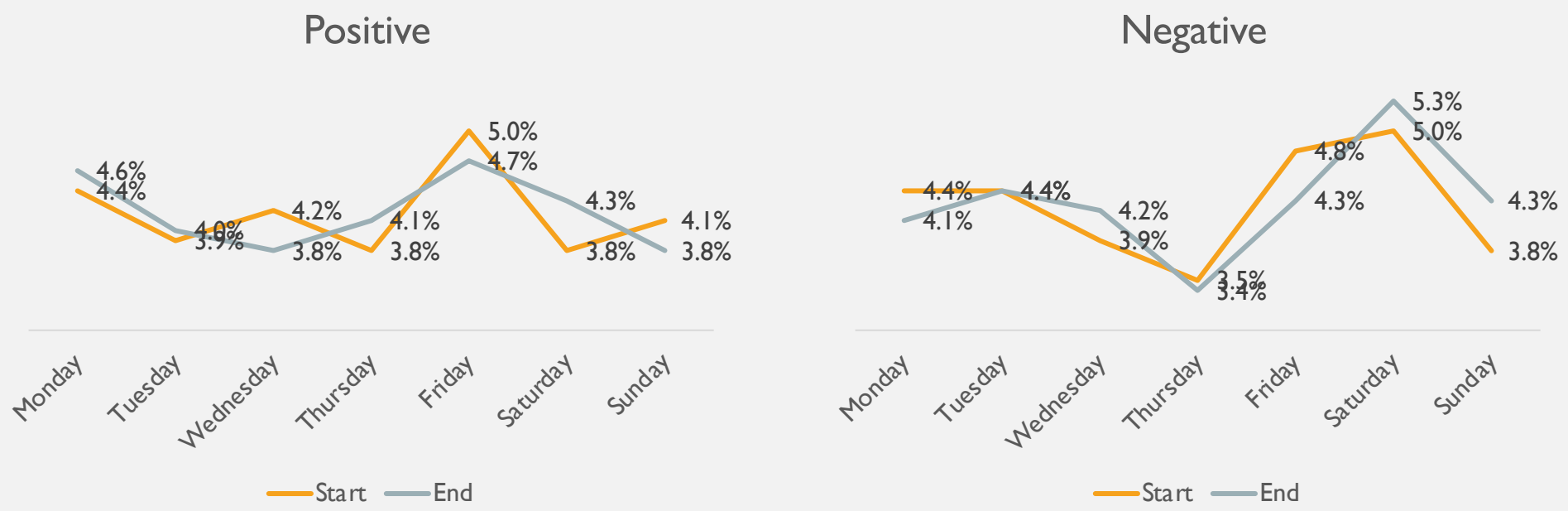
SHIFT STARTS ARE MOST PUNCTUAL ON FRIDAYS; MOST MONDAY SHIFTS DO NOT END ON TIME PLANNED



BASE: ALL SHIFTS, TOTAL DIFFERENCE

SHIFTS END LATEST ON FRIDAYS; SHIFTS ARE STARTED EARLIEST ON SATURDAYS

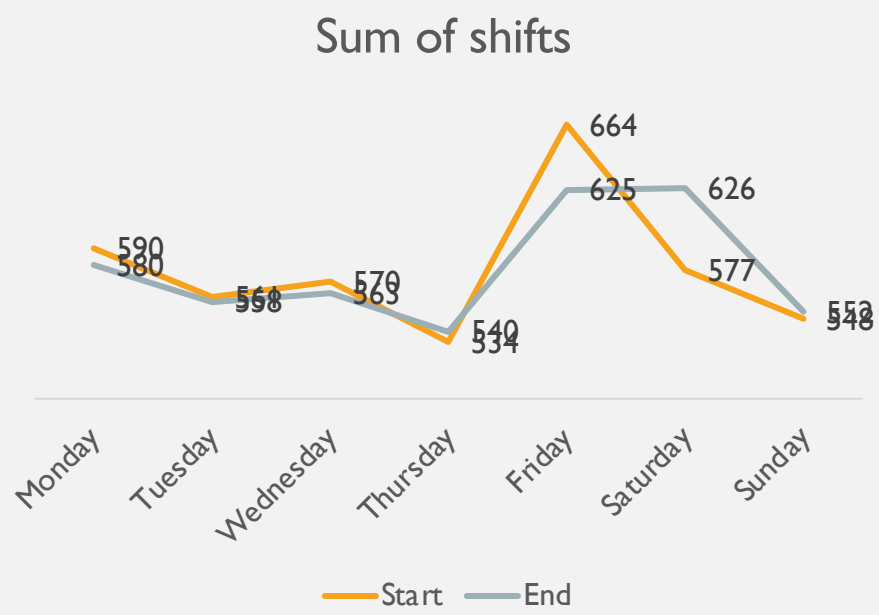
SHARE OF ALL SHIFTS ABOVE THRESHOLD BY WEEKDAY

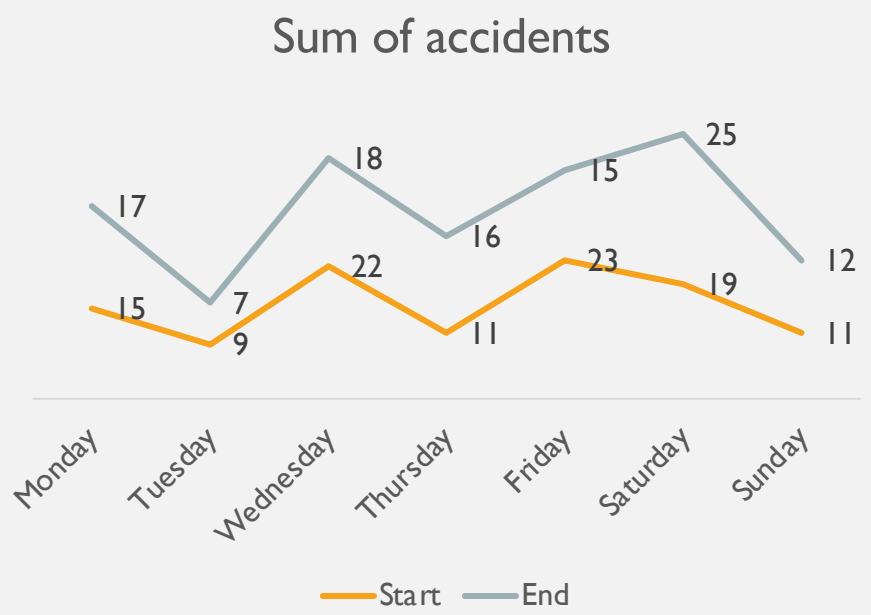


BASE: ALL SHIFTS, TOTAL DIFFERENCE, THRESHOLD: 5 MINS

MOST SHIFTS ARE ENDED ON FRIDAYS; MOST ACCIDENTS OCCUR WHEN SHIFTS END ON SATURDAYS

SUM OF SHIFTS AND ACCIDENTS BY WEEKDAY

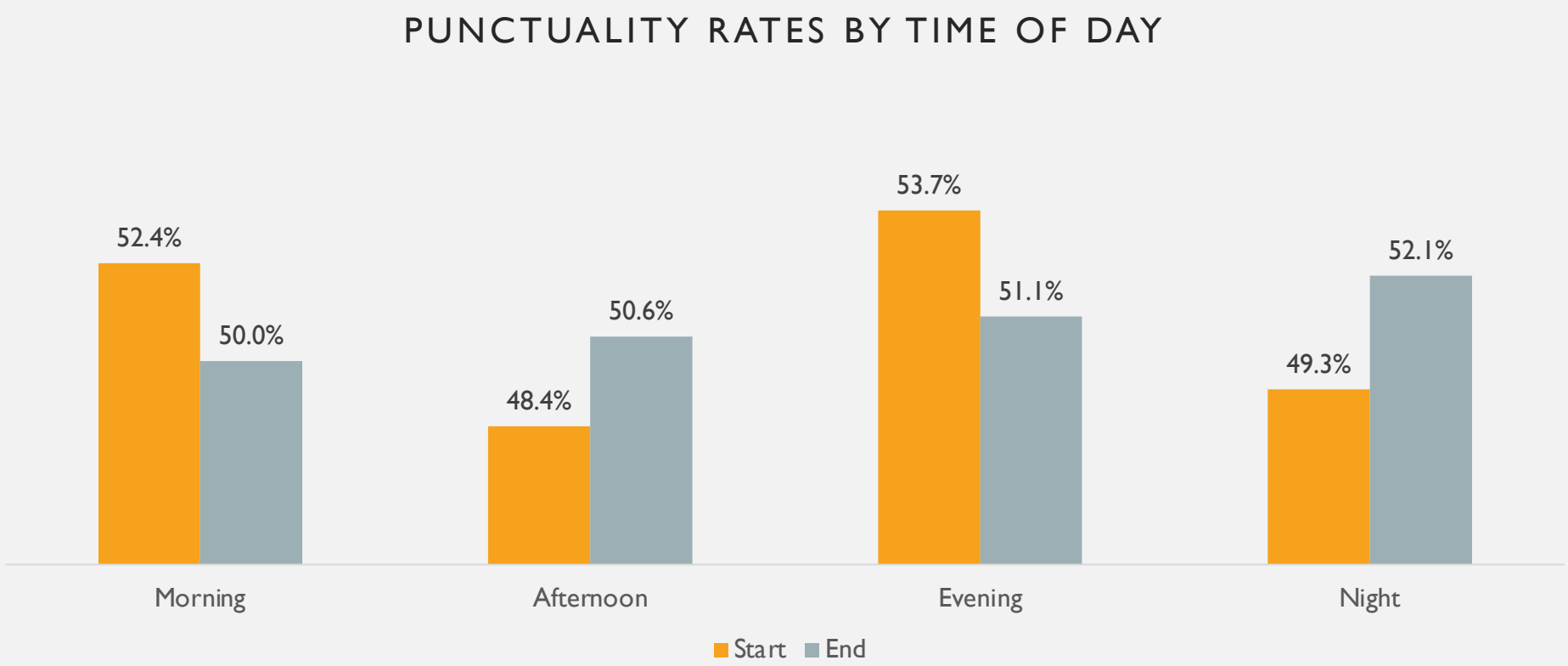




BASE: ALL SHIFTS

TIMES OF DAY

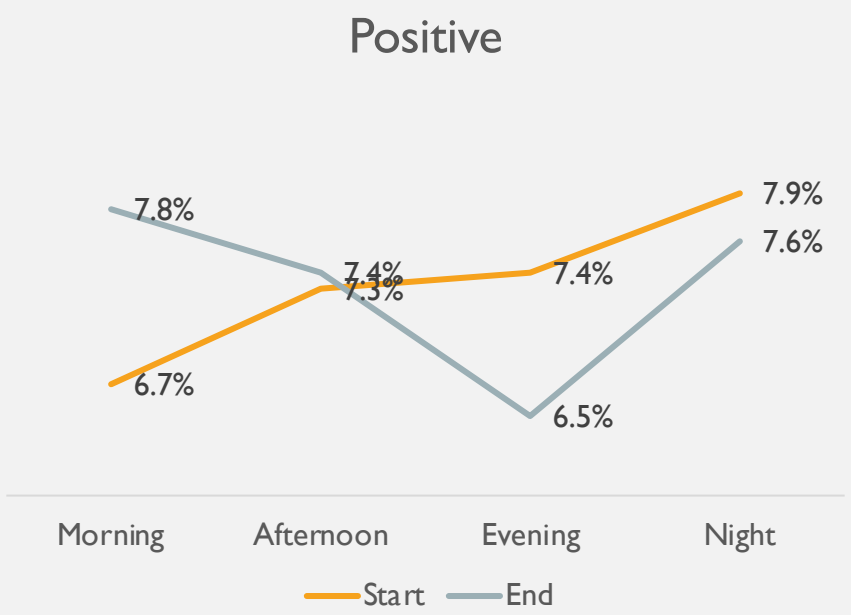
EVENING SHIFTS START THE MOST PUNCTUAL; MORE THAN EVERY SECOND NIGHT SHIFT IS ENDED ON TIME

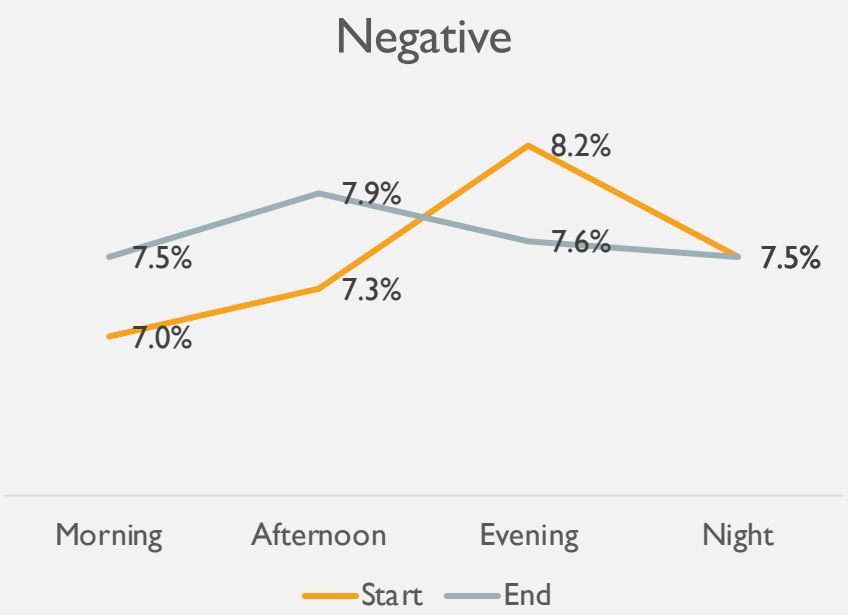


BASE: ALL SHIFTS, TOTAL DIFFERENCE

7.9% OF NIGHT SHIFTS ARE STARTED LATER; AFTERNOON SHIFTS ARE ENDED EARLIEST

SHARE OF ALL SHIFTS ABOVE THRESHOLD BY TIME OF DAY

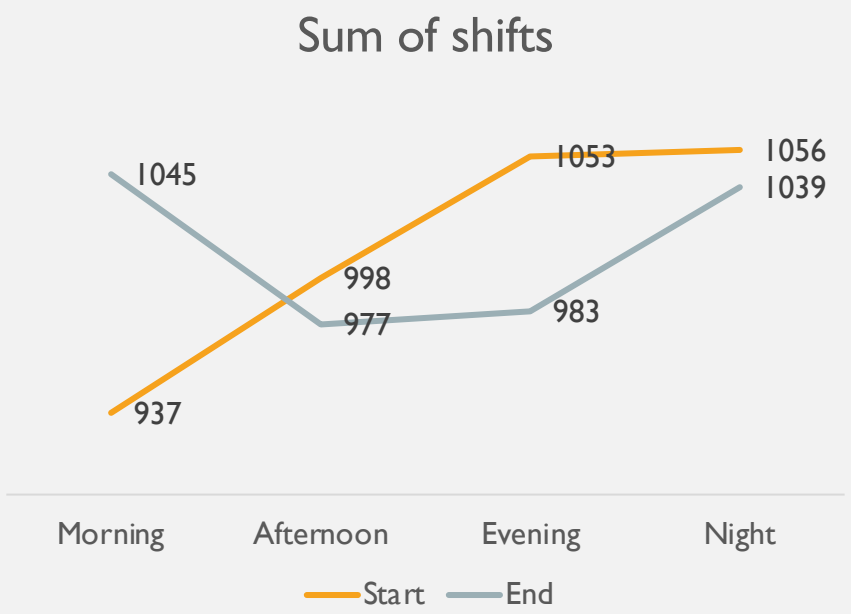


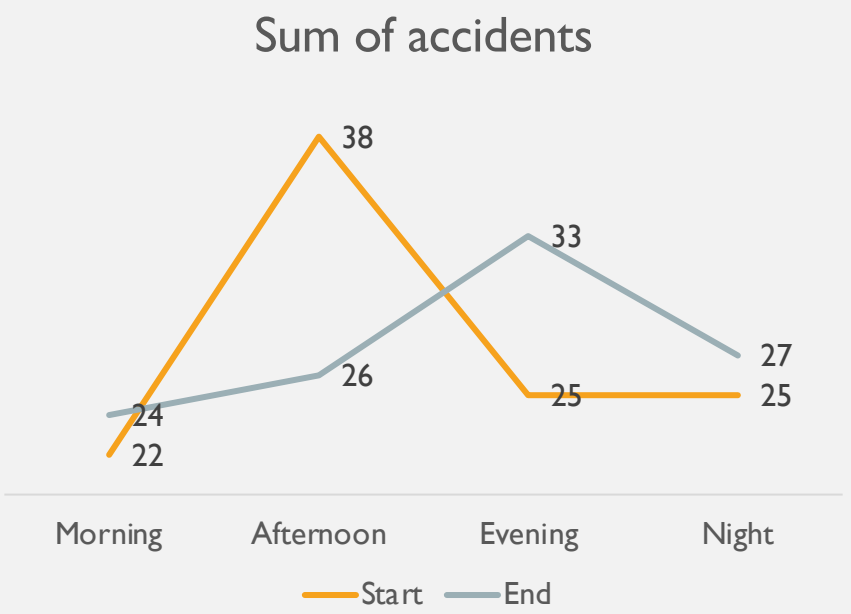


BASE: ALL SHIFTS, TOTAL DIFFERENCE, THRESHOLD: 5 MINS

MOST SHIFTS ARE STARTED AT NIGHT; MOST ACCIDENTS OCCUR WHEN SHIFT WAS STARTED IN THE AFTERNOON

SUM OF SHIFTS AND ACCIDENTS BY TIME OF DAY





BASE: ALL SHIFTS

DASHBOARD MOCKUP

Home | Comparison

Shifts | Areas | Drivers | Times of day | Weekdays

01/01/2021 — 15/02/2021 | Area | Driver | Rating | Threshold

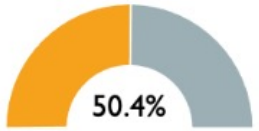
With accident

Saldo positive delay:
10.13 days

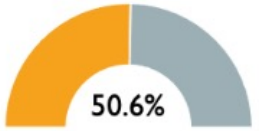
Saldo negative delay:
-10.53 days

Saldo total:
-9.58 hours

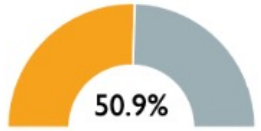
Punctuality Rate Start



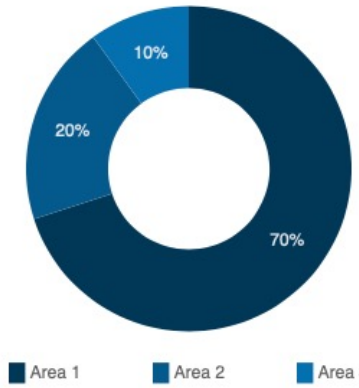
Punctuality Rate End



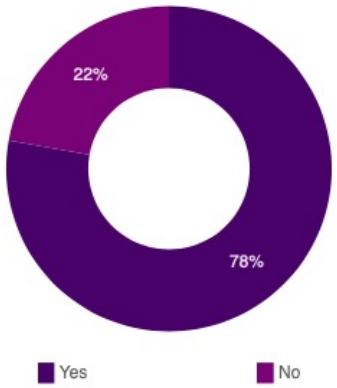
Punctuality Rate Total



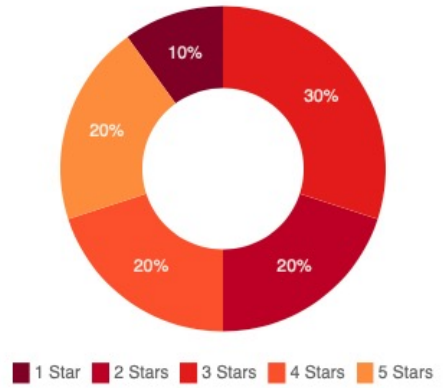
By area



By accident



By rating



TASK 2: INFLUENTIAL FACTORS

REGRESSION SHOWS THAT NONE OF THE AVAILABLE FACTORS HAS A SIGNIFICANT INFLUENCE ON A DRIVER'S PERFORMANCE

MULTIPLE LINEAR REGRESSION ANALYSIS

Total time difference (mean)

Independent variable	Coefficients	p-Values
has_safety_training	-0.019559	0.901625
driving_licence_issued_year	0.000107	0.916752
completed_shifts	-0.004661	0.799235
shift_length_avg_sec	0.000048	0.416005
accidents_sum	0.042021	0.920624
rating_avg_weighted	-0.057629	0.783769

Total time difference (sum)

Independent variable	Coefficients	p-Values
has_safety_training	-0.001340	0.850284
driving_licence_issued_year	-0.086110	0.541816
completed_shifts	-0.050936	0.582168
shift_length_avg_sec	0.000184	0.575213
accidents_sum	2.444349	0.314366
rating_avg_weighted	-0.457179	0.837305

Reading aid: For instance, for every one unit of change in the sum of accidents, the change in the performance (sum) is about 2.44%. In simpler words, if a driver has one more accident, they can expect to achieve an increase of 2.44% in their total time difference compared to before.

BASE: ALL DRIVERS

Reading aid: A low p-value (< 0.05) indicates that the null hypothesis can be rejected. In other words, a predictor that has a low p-value is likely to be a meaningful addition to a model because changes in the this value are related to changes in the response variable.

A DRIVER WITH THE PARAMETERS LISTED BELOW WILL HAVE AN AVERAGE TIME DIFFERENCE OF 0.16 MINUTES

MULTIPLE LINEAR REGRESSION ANALYSIS

Driver parameters

- has_safety_training: 1
- driving_licence_issued_year: 2010
- completed_shifts: 30
- shift_length_avg_sec: 20000.00
- accidents_sum: 1,
- rating_avg_weighted: 4.0

Predictions for this driver (total time diff)

Mean: **0.02** minutes

Sum: **3.78** minutes

BASE: ALL DRIVERS

DRIVERS THAT HAVE A TIME SALDO ABOVE THE POSITIVE THRESHOLD HAVE A SLIGHTLY LOWER ACCIDENT PROBABILITY

ACCIDENT PROBABILITY OF DRIVERS

Reading aid: The probability that drivers with a total time difference above the positive threshold had at least one accident is 8.2% compared to 9.2% for drivers with a total time difference below the positive threshold. Hence, it is more probably for the second group that his specific event (an accident) happens.

Above positive threshold:
8.2%

vs.

Below positive threshold:
9.2%

Above negative threshold:
9.2%

vs.

Below negative threshold:
9.1%

BASE: ALL DRIVERS, TOTAL DIFFERENCE, THRESHOLD: 5 MINS

DRIVERS THAT WORK LESS ACCORDING TO THE THRESHOLD ARE MUCH MORE PROBABLY TO BE RATED BADLY

PERFORMANCE PROBABILITY OF DRIVERS

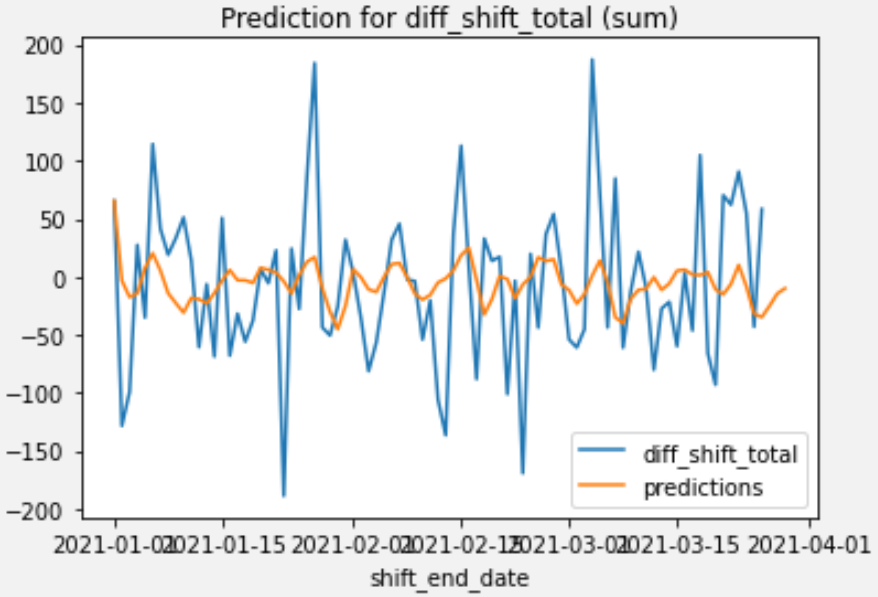
Bad		Okay		Good	
Above positive threshold:	Above negative threshold:	Above positive threshold:	Above negative threshold:	Above positive threshold:	Above negative threshold:
677.9%	0%	126.5%	112.5%	79.9%	94.0%
vs.	vs.	vs.	vs.	vs.	vs.
Below positive threshold:	Below negative threshold:	Below positive threshold:	Below negative threshold:	Below positive threshold:	Below negative threshold:
28.1%	115.6%	96.7%	98.0%	102.5%	100.9%

BASE: ALL DRIVERS, TOTAL DIFFERENCE, THRESHOLD: 5 MINS

EXTRA: FORECAST

FORECAST SHOWS THAT TOTAL TIME DIFFERENCE OF ALL SHIFTS WILL BE NEGATIVE IN THE FOLLOWING THREE DAYS

ARIMA FORECAST (1/2)

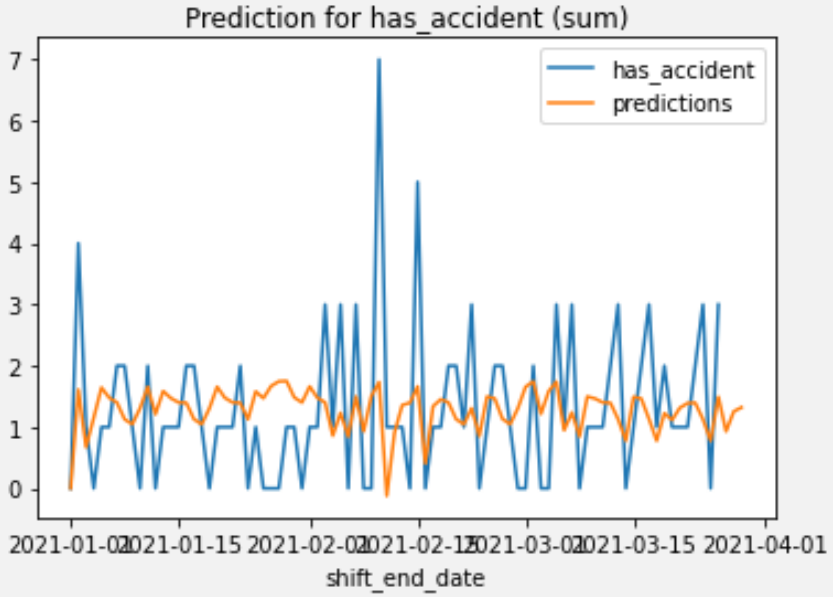


shift_end_date	
2021-01-01	65.379097
2021-01-02	-3.334781
2021-01-03	-17.553458
2021-01-04	-14.322372
2021-01-05	7.480108
	...
2021-03-25	-32.109370
2021-03-26	-34.799377
2021-03-27	-25.155026
2021-03-28	-14.564991
2021-03-29	-10.051764

BASE: ALL SHIFTS; PREDICTION FOR FOLLOWING THREE DAYS AFTER LATEST DATE AVAILABLE IN DATASET

FORECAST INDICATES THAT ACCIDENT PROBABILITY WILL INCREASE WITHIN THE FOLLOWING THREE DAYS

ARIMA FORECAST (2/2)



shift_end_date	
2021-01-01	0.000000
2021-01-02	1.610962
2021-01-03	0.672380
2021-01-04	1.154612
2021-01-05	1.644709
	...
2021-03-25	0.777877
2021-03-26	1.492304
2021-03-27	0.930283
2021-03-28	1.255638
2021-03-29	1.320541

BASE: ALL SHIFTS; PREDICTION FOR FOLLOWING THREE DAYS AFTER LATEST DATE AVAILABLE IN DATASET

METHODOLOGY

METHODOLOGY (1/8)

Time difference

The **start** difference is calculated by substracting the planned start timestamp from the actual start one.

The **end** difference is calculated by substracting the planned end timestamp from the actual end one.

The **total** difference is the result of adding up the start and end differences.

Punctuality Rates

The **start** rate is calculated by dividing the number of shifts that have a start diff equal to or lower 0 with the number of all shifts.

The **end** rate is calculated by dividing the number of shifts that have an end diff equal to or lower 0 with the number of all shifts.

The **total** rate is calculated by dividing the number of shifts that have a total diff equal to or lower 0 with the number of all shifts.

METHODOLOGY (2/8)

Percentiles

Percentiles help to understand the relative standing of a value and identify where the value falls within a distribution of values. They tell you how a value compares to other values. The general rule is that if value X is at the kth percentile, then X is greater than K% of the values.

The first quartile, also known as Q1 or the lower quartile, is the value of the 25th percentile. The bottom quarter of the scores fall below this value, while three-quarters fall above it.

The second quartile, also known as Q2 or the median, is the value of the 50th percentile. Half the scores are above and half below.

The third quartile, also known as Q3 or the upper quartile, is the value of the 75% percentile. The top quarter of the scores fall above this value, while three-quarters fall below it.

Standard deviation

In statistics, the standard deviation is a measure of the amount of variation or dispersion of a set of values. A low standard deviation indicates that the values tend to be close to the mean (also called the expected value) of the set, while a high standard deviation indicates that the values are spread out over a wider range.

Standard deviation may be abbreviated SD or std, and is most commonly represented in mathematical texts and equations by the lower case Greek letter sigma σ, for the population standard deviation, or the Latin letter s, for the sample standard deviation.

METHODOLOGY (3/8)

Threshold

The threshold is a variable that can be set in the Python script that builds the foundation for this analysis. The idea of the threshold is to separate shifts, drivers etc. into relevant and irrelvant subsets. This helps focusing on data points that represent problematic developments within the operations business as differences or delays below a certain threshold are normal and will never be eliminated.

Rating weightings

Ratings of drivers with more frequent customer evaluations are more valuable. In order to include this frequency, the number of ratings per driver is cut into five categories which represent the weightings (from 1.1 to 1.5). The average rating of each driver is eventually multiplied with the respective weighting to get a weighted average rating for each driver.

Due to the maximal weight of 1.5 the maximum rating of 5 will become 7.5.

METHODOLOGY (4/8)

Rating categories

By categorizing the rating data we can get insights on our drivers that allow us to understand better if drivers with certain customer rating levels behave differently. For example, one assumption here would be that the better a driver is rated in total, the more they work.

Bad: avg. weighted rating >= 0 and < 2.5

Okay: avg. weighted rating >= 2.5 and < 5.0

Good: avg. weighted rating >= 5.0 and 7.5

Skill levels

By categorizing the completed shifts data we can get insights on our drivers that allow us to understand better if drivers with certain experience levels act differently. For example, one assumption here would be that the more experienced a driver is, the less accidents they cause.

Beginner: 0 to 49 completed shifts

Advanced: 50 to 99 completed shifts

Export: 100+ completed shifts

METHODOLOGY (5/8)

Time of days

By categorizing the time data we can get insights for specific times of the day that allow us to understand better if certain hours of the day show special developments. For example, one assumption here would be that the later the day a shift ends, the more probable accidents become.

Night: hour >= 0 and < 6

Morning: hour >= 6 and < 12

Afternoon: hour >= 12 and < 18

Evening: hour >= 18 and < 24

METHODOLOGY (6/8)

Multiple linear regression

Multiple linear regression is used to estimate the relationship between two or more independent variables and one dependent variable. You can use multiple linear regression when you want to know:

- How strong the relationship is between two or more independent variables and one dependent variable (e.g. how rainfall, temperature, and amount of fertilizer added affect crop growth).
- The value of the dependent variable at a certain value of the independent variables (e.g. the expected yield of a crop at certain levels of rainfall, temperature, and fertilizer addition).

Coefficients

Regression coefficients are estimates of the unknown population parameters and describe the relationship between a predictor variable and the response. In linear regression, coefficients are the values that multiply the predictor values. Suppose you have the following regression equation: $y = 3X + 5$. In this equation, +3 is the coefficient, X is the predictor, and +5 is the constant.

The sign of each coefficient indicates the direction of the relationship between a predictor variable and the response variable.

- A positive sign indicates that as the predictor variable increases, the response variable also increases.
- A negative sign indicates that as the predictor variable increases, the response variable decreases.

The coefficient value represents the mean change in the response given a one unit change in the predictor. For example, if a coefficient is +3, the mean response value increases by 3 for every one unit change in the predictor.

METHODOLOGY (7/8)

p-Value

The p-value for each term tests the null hypothesis that the coefficient is equal to zero (no effect). A low p-value (< 0.05) indicates that you can reject the null hypothesis. In other words, a predictor that has a low p-value is likely to be a meaningful addition to your model because changes in the predictor's value are related to changes in the response variable.

Conversely, a larger (insignificant) p-value suggests that changes in the predictor are not associated with changes in the response.

R2

R-squared is a goodness-of-fit measure for linear regression models. This statistic indicates the percentage of the variance in the dependent variable that the independent variables explain collectively. R-squared measures the strength of the relationship between your model and the dependent variable on a convenient 0 – 100% scale.

Adjusted R-squared can provide a more precise view of that correlation by also taking into account how many independent variables are added to a particular model against which the stock index is measured. This is done because such additions of independent variables usually increase the reliability of that model—meaning, for investors, the correlation with the index.

METHODOLOGY (8/8)

Forecast

An ARIMA model is a class of statistical models for analyzing and forecasting time series data. It explicitly caters to a suite of standard structures in time series data, and as such provides a simple yet powerful method for making skillful time series forecasts.

ARIMA is an acronym that stands for AutoRegressive Integrated Moving Average. It is a generalization of the simpler AutoRegressive Moving Average and adds the notion of integration. This acronym is descriptive, capturing the key aspects of the model itself. Briefly, they are:

- AR: Autoregression. A model that uses the dependent relationship between an observation and some number of lagged observations.
- I: Integrated. The use of differencing of raw observations (e.g. subtracting an observation from an observation at the previous time step) in order to make the time series stationary.
- MA: Moving Average. A model that uses the dependency between an observation and a residual error from a moving average model applied to lagged observations.

Each of these components are explicitly specified in the model as a parameter. A standard notation is used of ARIMA (p,d,q) where the parameters are substituted with integer values to quickly indicate the specific ARIMA model being used.

The parameters of the ARIMA model are defined as follows:

- p: The number of lag observations included in the model, also called the lag order.
- d: The number of times that the raw observations are differenced, also called the degree of differencing.
- q: The size of the moving average window, also called the order of moving average.

A linear regression model is constructed including the specified number and type of terms, and the data is prepared by a degree of differencing in order to make it stationary, i.e. to remove trend and seasonal structures that negatively affect the regression model.

A value of 0 can be used for a parameter, which indicates to not use that element of the model. This way, the ARIMA model can be configured to perform the function of an ARMA model, and even a simple AR, I, or MA model.

Adopting an ARIMA model for a time series assumes that the underlying process that generated the observations is an ARIMA process. This may seem obvious, but helps to motivate the need to confirm the assumptions of the model in the raw observations and in the residual errors of forecasts from the model.