

Data Analyst Case Work Presentation

CarCompany

Rezso Roland Gimesi

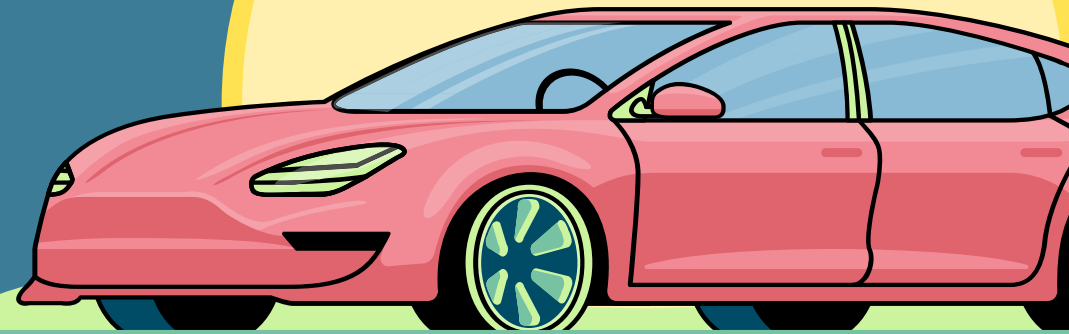


Table of content

1. Introduction to the case

2. Requested insights

3. Data analysis tools

4. Analysis in detail

- How I achieved the results
- Results

5. **Suggestions**

The Case

CarCompany's core service is a auction platform for trading second-hand vehicles between dealerships.

Guaranteed Price service: estimates the price a car can be sold on the platform.

Takes over the risk from clients as CarCompany covers the difference if a car can only be sold below the guaranteed price.

Due to the high risk of the service, continuous evaluation of performance is required.





Requested Insights

01

Distribution of GP accuracy

What is the distribution where GP is higher than the sold price, equal to the sold price, and lower than the sold price?

03

Employee performance

What the ranking of employees based on acceptance rate? What more can we look into in relation to acceptance of guaranteed prices?

02


GP acceptance rate

What percentage of provided GP's are accepted?

04

GP– Sold price correlation

Is there a correlation between the guaranteed price and the sold price?





Toolkit **overview**

Python

Data manipulation,
Analysis of insights 1-2-4

SQL

Data manipulation,
Analysis of insight 3

Power BI

Data visualization



1. Distribution of GP accuracy – analysis

What is the distribution where GP is higher than the sold price, equal to the sold price, and lower than the sold price?

Tools: Python, Power BI

1. Imported data to Jupyter Notebook

2. Basic standardization

3. Added columns

- Deviation from GP
- GP vs SP difference

5. Distribution check

6. Save for viz and further analysis

```
jupyter AUToproff case work Last Checkpoint: 23 minutes ago (autosaved)
File Edit View Insert Cell Kernel Widgets Help Not Trusted

In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# Import the data
# Set the thousand separator to '.' and the decimal separator to ',' (prevents recognizing data as string)
# keep employee id (guaranteed_price_provider) as object
cars_df = pd.read_csv('C:/Users/grezo/Desktop/Rencool/Copenhagen/Job hunt/trial work/AUToproff/case_data.csv',
dtype={'guaranteed_price_provider': 'object'}, thousands='.', decimal=',')

In [2]: # Convert everything to integer so all numerical data have the same data type
cars_df['guarantee_bid_price (DKK)'] = cars_df['guarantee_bid_price (DKK)'].astype('int64')

# Impute null values. No info on reason, so we will use 0 for convenience.
cars_df['minimum_price'] = cars_df['minimum_price'].fillna(0)
cars_df['minimum_price'] = cars_df['minimum_price'].astype('int64')

In [3]: # Create a column to represent the deviance of the Selling Price from the Guaranteed Price
cars_df['GP_difference'] = cars_df['sold_price'] - cars_df['guarantee_bid_price (DKK)']

# Add a column to indicate if GP or SP is higher (or equal)
cars_df['GP_vs_SP'] = ''

for i, row in cars_df.iterrows():
    val = row['GP_difference']

    if val > 0:
        cars_df.at[i, 'GP_vs_SP'] = 'SP higher'
    elif val == 0:
        cars_df.at[i, 'GP_vs_SP'] = 'equal'
    elif val < 0:
        cars_df.at[i, 'GP_vs_SP'] = 'GP higher'

In [4]: # Distribution of GP vs SP
cars_df['GP_vs_SP'].value_counts(normalize=True)

Out[4]: SP higher    0.808731
equal          0.106784
GP higher       0.084485
Name: GP_vs_SP, dtype: float64
```

Python analysis
Full notebook available [here](#).

1. Distribution of GP accuracy – results

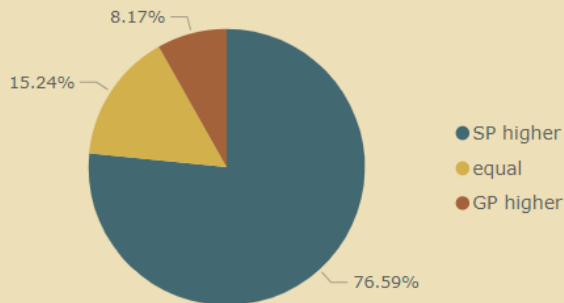
What is the distribution where GP is higher than the sold price, equal to the sold price, and lower than the sold price?

- **More than 91% accurate overall** (when SP is equal or higher than GP)
- **Other interesting splits?**
 - **How the distribution changes between accepted/declined offers**

GP offer status

- Accepted
- Declined

Distribution of GP accuracy

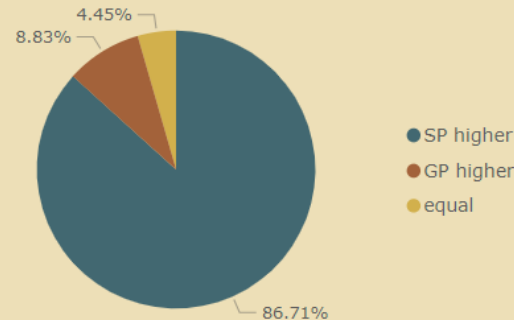


Accepted offers

GP offer status

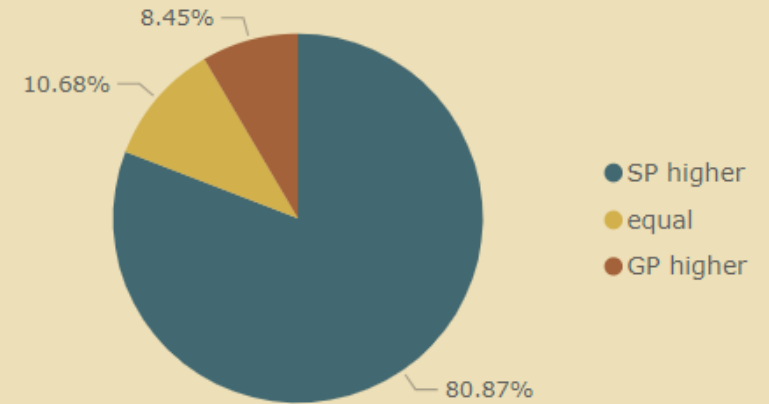
- Accepted
- Declined

Distribution of GP accuracy



Declined offers

Distribution of GP accuracy



Total accuracy

SP is higher in more than 10% of the times in declined cases.

Clients right to assume that they can have a higher price

2. GP acceptance rate

What percentage of provided GP's are accepted?

Tools: Power BI

- Data already provided (guarantee_bid_status)
- Only needed to visualize
- Slightly over 57% of the offers get accepted.

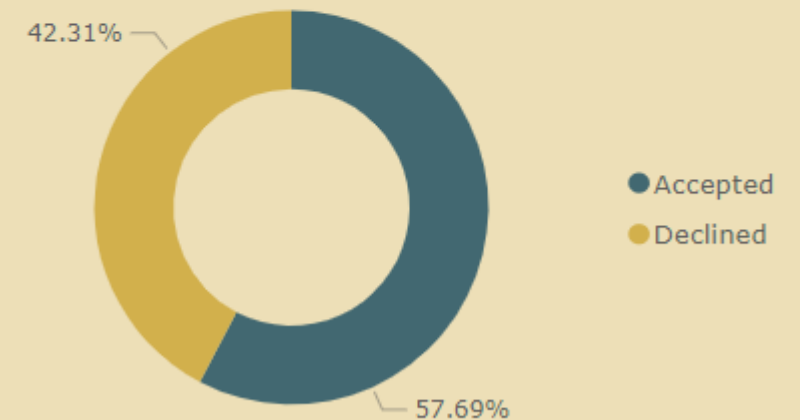
How to analyze this further:

How does the acceptance rate change over time?
Potential impact of age/use of the vehicle?

GP offers made:

3184

GP acceptance rate



GP acceptance rate

3. Employee ranking – analysis

What the ranking of employees based on acceptance rate? What more can we look into in relation to acceptance of guaranteed prices?

Tools: SQL, Power BI

1. Imported data to SSMS
2. Created table for employee rank data
3. Created counts for accepted and declined GP offers for every employee
4. Calculated the acceptance ratio
5. Created view for viz

```
SQLQuery2.sql - DE...587P8E\grezs (57)* X
select * from case_data_manipulated$

drop table if exists employee_rank
create table employee_rank (
    guaranteed_price_provider nvarchar(50),
    accepted numeric,
    declined numeric
)

insert into employee_rank
select guaranteed_price_provider,
count(case when guarantee_bid_status = 'Accepted' then 1 end) as accepted,
count(case when guarantee_bid_status = 'Declined' then 1 end) as declined
from case_data_manipulated$
group by guaranteed_price_provider

drop view if exists employee_rank_view
go
create view employee_rank_view as
select *, cast(accepted/(accepted+declined) as decimal(7,2)) as acc_ratio
from employee_rank
--order by 3
```

SQL query

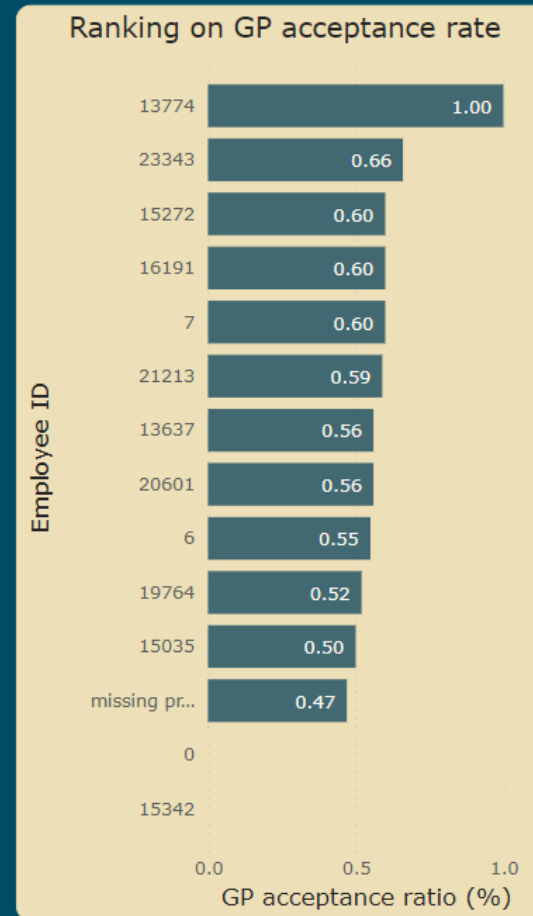
3. Employee ranking – results

What the ranking of employees based on acceptance rate? What more can we look into in relation to acceptance of guaranteed prices?

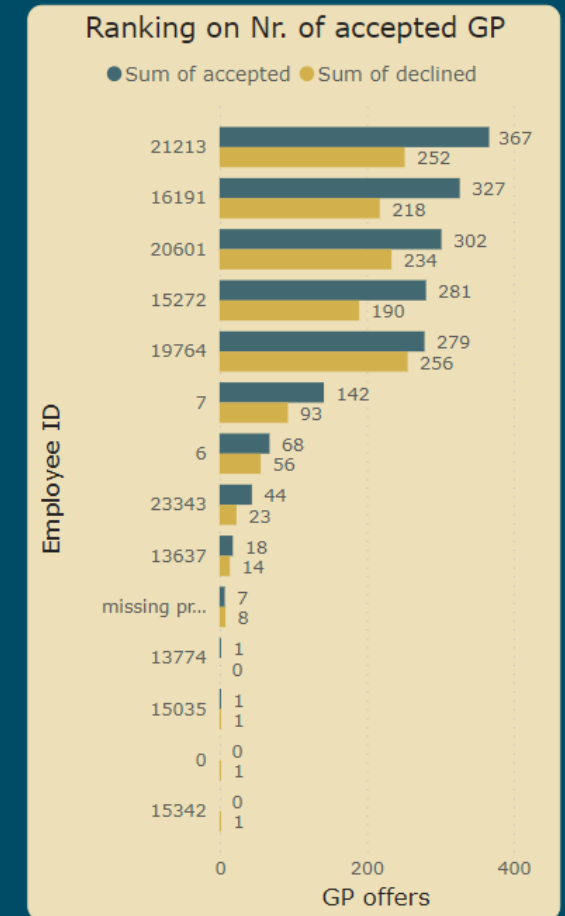
- **Employee ranking based on acceptance rate**
- **However, not a sufficient measure** (don't account for the amount of offers sent)
- **Should include the amount of accepted offers sent in the ranking**

How to analyze this further:

How does performance change over time?
What percent of the employee's accepted offers turn out to be lossmaking?



Ranking on acceptance rate



Ranking on accepted offers

4. GP – Sold Price correlation

Is there a correlation between the guaranteed price and the sold price?

Tools: Python, Power BI

Assumption: there is correlation (lets test it)

1. Plotted the correlation matrix on a heat map

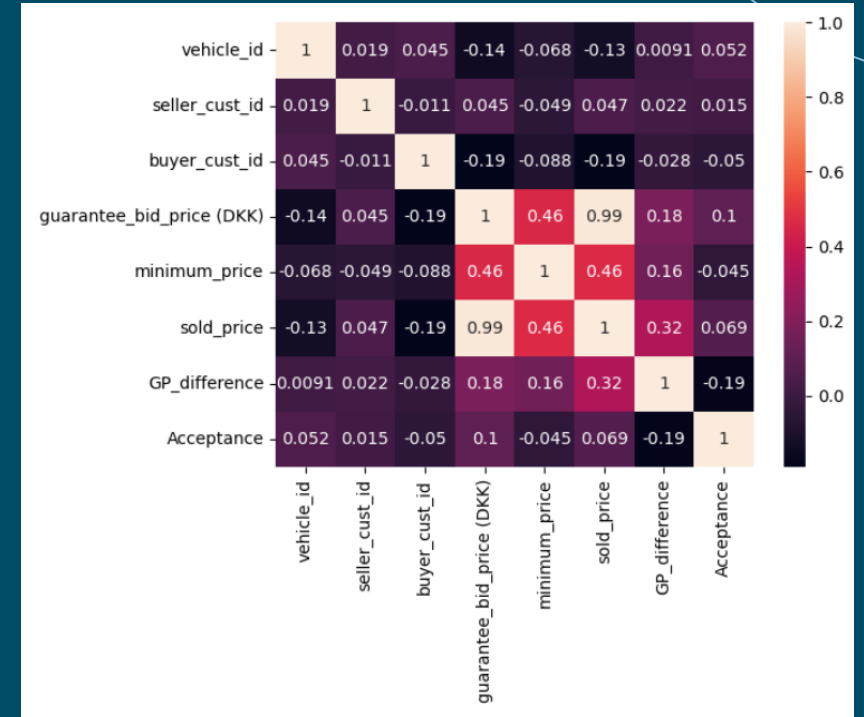
- Seems like there's a strong correlation

2. Tested the correlation of the two variables

- $\text{Corr.} = 0.988$, $p = 0.0$



Regression line on offered GP – SP relationship

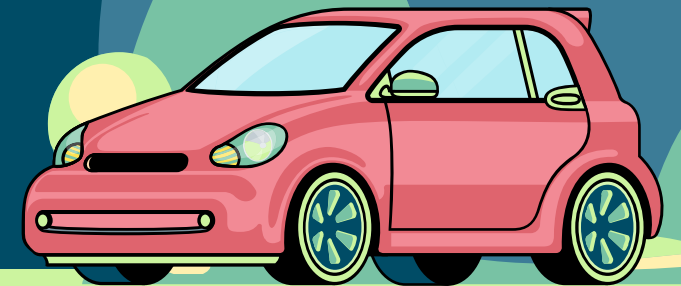


Correlation heat map

Based on the tests above, there is a strong, positive relationship between the 2 variables, which is statistically significant

Suggestions

1. CarCompany should investigate how does the acceptance rate change over time.
2. CarCompany should investigate how does the employee performance change over time. Moreover, they need to introduce other employee ranking metrics, that weights for the amount of accepted offers that are profitable.
3. CarCompany should investigate other (not explicit) variables that might have an impact on acceptance rate (age, usage, etc.)



Thank you!

Do you have any questions?

