



EDA Project “Muesli”

Contributors

Alice Fritze

Isamin Gubian

Mark Jäger

Description	Page
Content	2
Mission	3
Planning	4
Process description/ KPIs	5
Overview	6
KPI 1	7
KPI 2	9
KPI 3	12
KPI 4	13
Further findings to consider	14
Conclusion	15

Tuesday

Prepare Planning

- Tasks/duration

Understand Business Problem and Potential Solution

- Map process (flow chart)
- Invent KPIs for performance (overall/stages) tracking
- Label flow chart with logic/assumptions of each stage

Explore Data

- Prepare/export four data sets (excel or csv)
- Load data into pandas and perform simple EDA
- Identify missing values and outliers if applicable
- Create new columns to represent processing time (incl. logic needed to handle weekends of processing modes etc.)
- Create new data frames by joining tables

Wednesday

Wednesday

Validate Solution

- Make further EDA to understand the time at each step and range of values for each KPI
- Identify problematic data or outliers if appl.
- Show average duration of order
- Show variation for each stage
- Evaluate alignment between data and assumptions
- Identify steps with 'concerning' levels of reliability

Thursday

Visualise and Communicate

- Prepare Jupyter notebook
- Prepare presentation

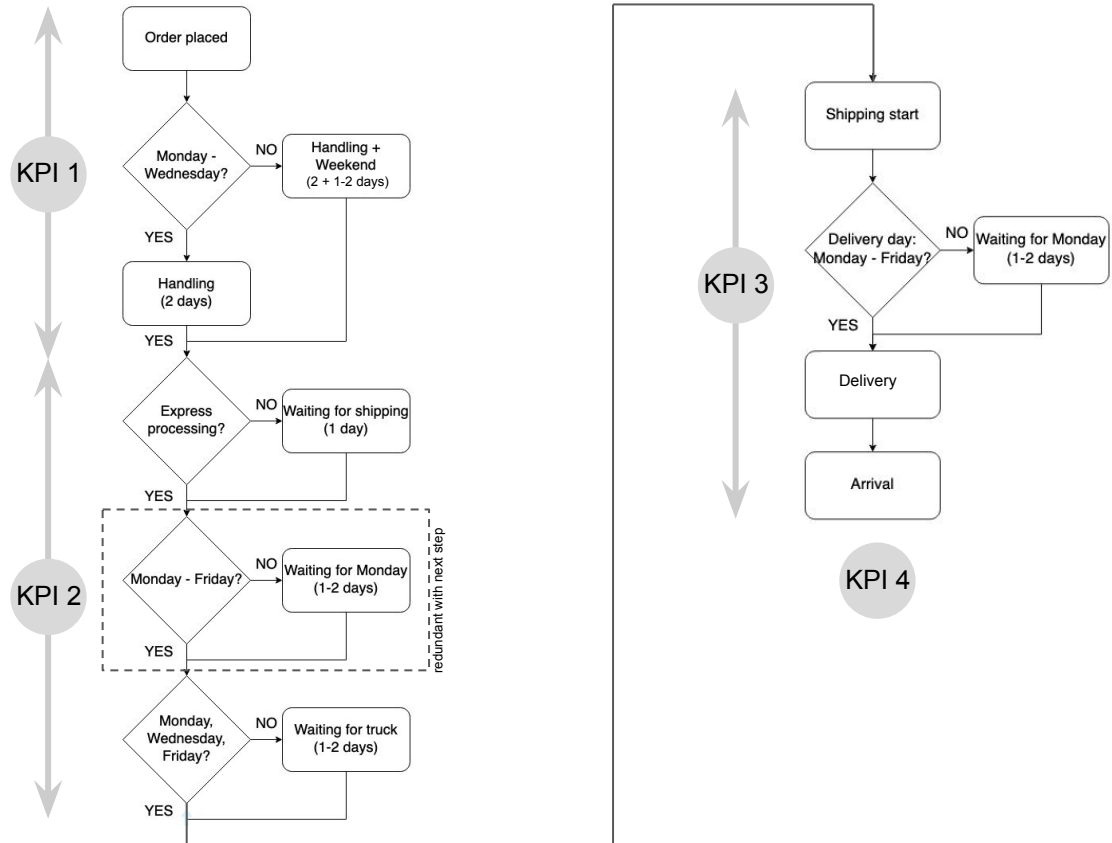
“Develop KPIs to help us keep track of the health of our business in order to improve the service we offer to our customers.”

KPIs for main process phases

- ❑ KPI 1 Duration in warehouse
- ❑ KPI 2 Duration between warehouse and shipping
- ❑ KPI 3 Duration of shipping
- ❑ KPI 4 Duration of overall process

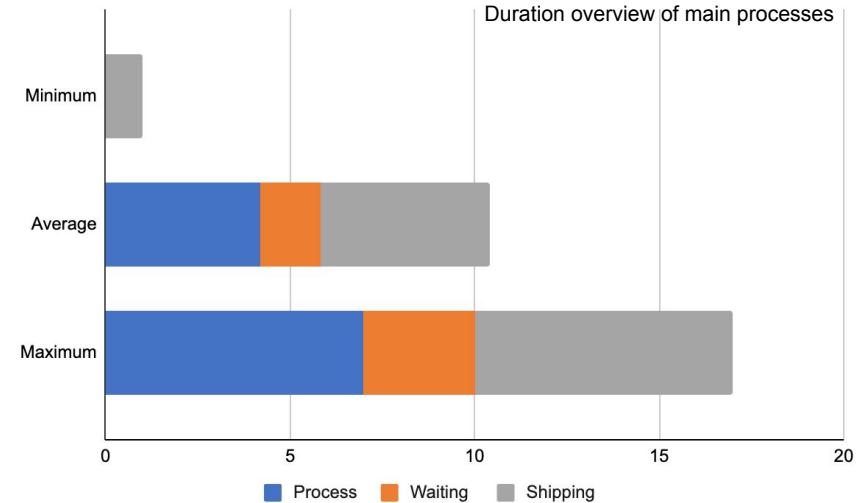
Reasons for KPIs

The chosen KPIs reflect the main parts of the process and need to be investigated prior to any further exploration.



Main observations

- ❑ Average warehouse processing was ~50% longer than expected
- ❑ Orders placed on Sunday require a ~2 day longer warehouse processing period
- ❑ The warehouse processing duration shows no reliability in regard to the estimated time
- ❑ Warehouse processing time without express is nearly three days longer than warehouse processing time with express
- ❑ Average duration between warehouse and shipping was ~80% longer than expected
- ❑ Average shipping duration was ~50% longer than expected
- ❑ There is a difference from minimum to maximum overall delivery time of 16 days



KPI 1 - Duration in warehouse

Description

Time between the time of the order and the finalization of the order in the warehouse

Assumption [days]

2 average

Data frame

Muesli Project raw data.xlsx/ InternDataStudy

Muesli Project raw data.xlsx/ Order Process Data

205 extracted samples

Main descriptions of data [days]

4.176 average

4 median

5 mode

0 minimum

8 maximum

1.97 std

Main observations

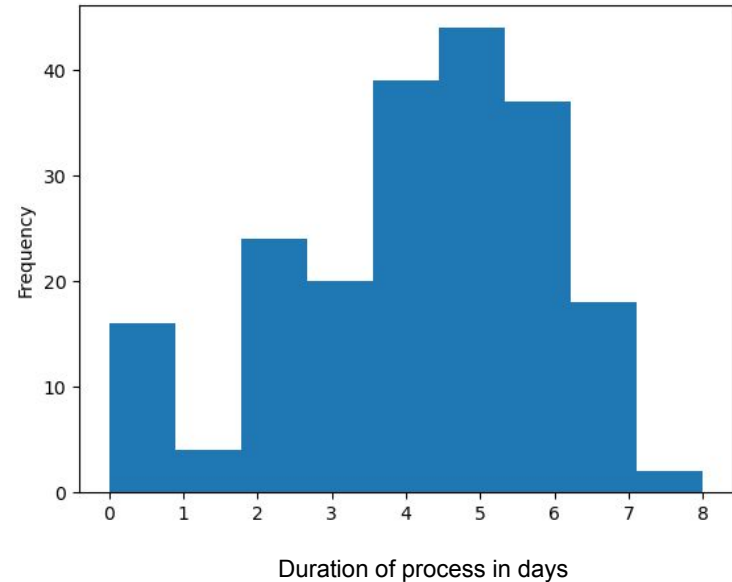
- ❑ Average duration was ~50% longer than expected

Warehouse process time without express

4.927

Warehouse process time with express

2.038



KPI 1 - Duration in warehouse and duration of waiting time

Description

Row1: Time between the time of the order and the finalization of the order in the warehouse
Row2: Waiting time for the order to be picked up by logistics

Data frame

Muesli Project raw data.xlsx/ InternDataStudy
Muesli Project raw data.xlsx/ Order Process Data
205 extracted samples

	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Process time	3.77	4.31	4.13	3.87	4.54	4.00	6.67
Waiting time for pickup	2.35	0	2.03	2.35	2.32	N/A	N/A

Main observations

- ❑ Orders placed on Sunday require a 2 day longer warehouse processing period

KPI 2 - Duration between warehouse and shipping

Description

Time between the order is ready in the warehouse to the moment it gets picked up by logistics company

Assumption

Monday, Wednesday and Friday (every other day)

Data frame

Muesli Project raw data.xlsx/ InternDataStudy

Muesli Project raw data.xlsx/ Order Process Data

290 extracted samples

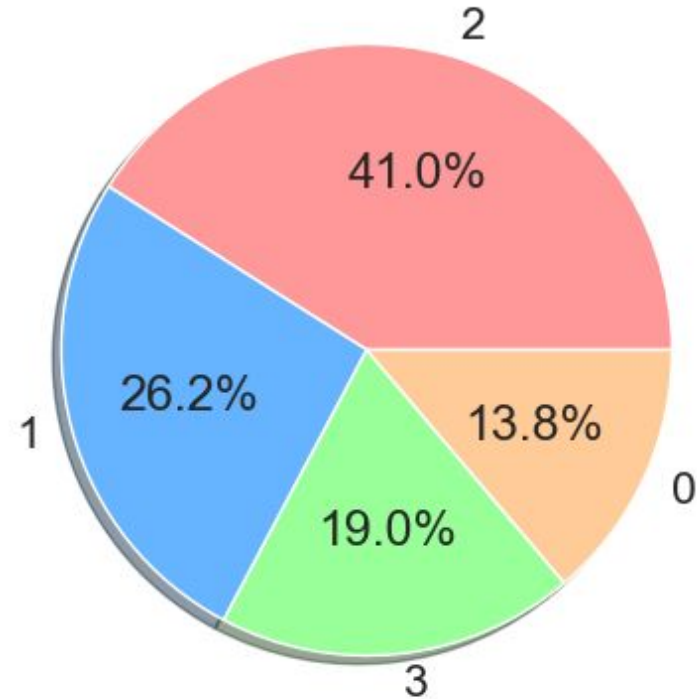
Main descriptions of data [days]

1.65 mean
2 median
2 mode
0 minimum
3 maximum

Main observations

- Waiting time is approximately 80% longer

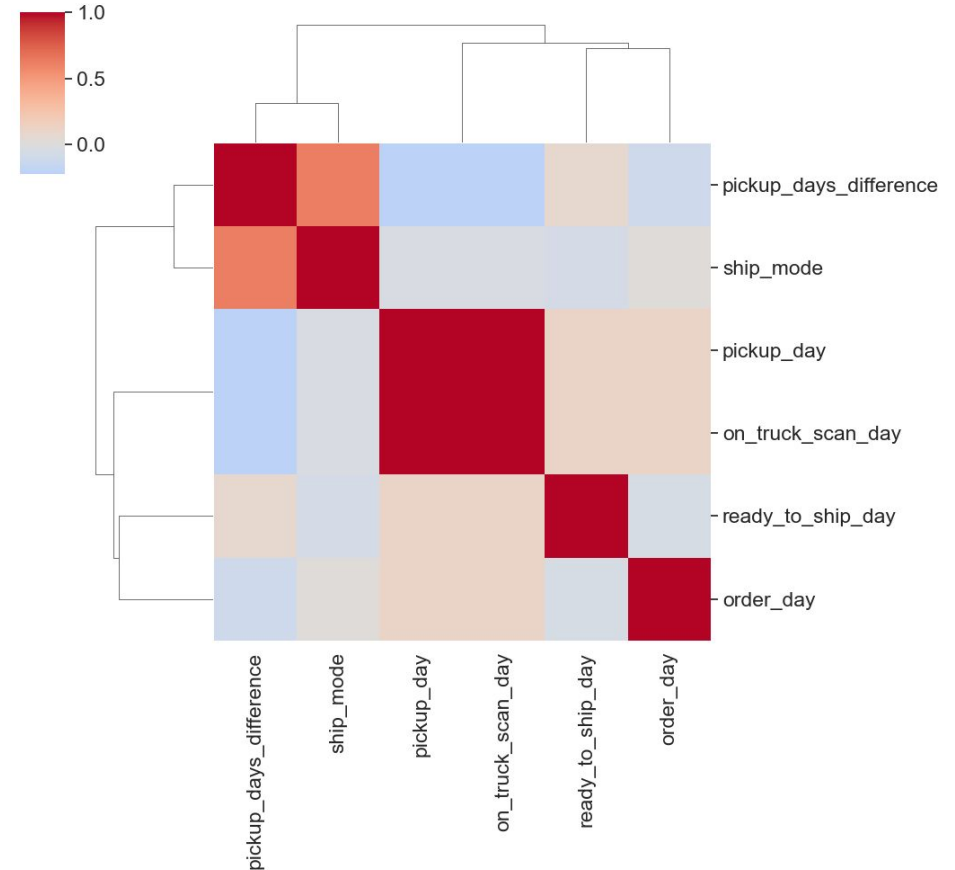
Waiting days for order pickup by logistics



KPI 2 - Duration between warehouse and shipping

Description

Heatmap used to investigate other possible correlations



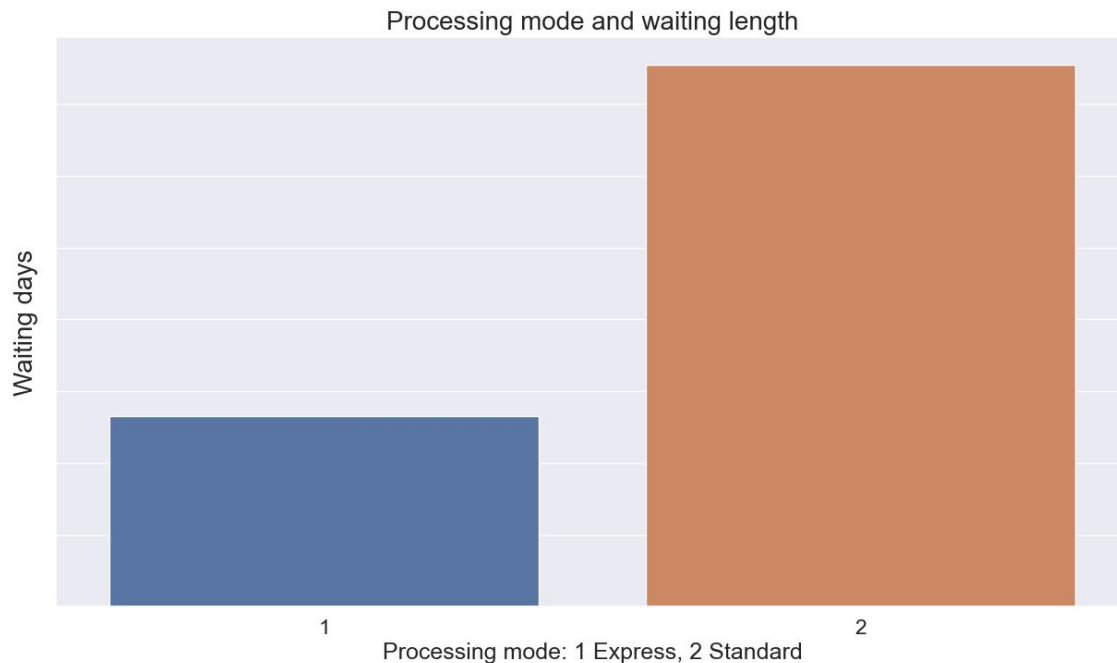
Main observations

- ❑ Correlation between waiting time and processing mode
- ❑ There is no correlation between day of the week and waiting days

KPI 2 - Duration between warehouse and shipping

Hypothesis

There is a correlation between the waiting period for logistics pickup and the processing mode chosen by the customer.



Main observations

- ❑ There is a positive correlation (calculated value = 0.721)

```
data1_df["pickup_days_difference"].corr(data1_df["ship_mode"])
```

KPI 3 - Duration of shipping

Description

Overall time used for the shipping process

Assumption [days]

3 average

Data frame

Muesli Project raw data.xlsx/ Campaign Data

Muesli Project raw data.xlsx/ Order Process Data

333 extracted samples

Main descriptions of data [days]

4.60 average

5 median

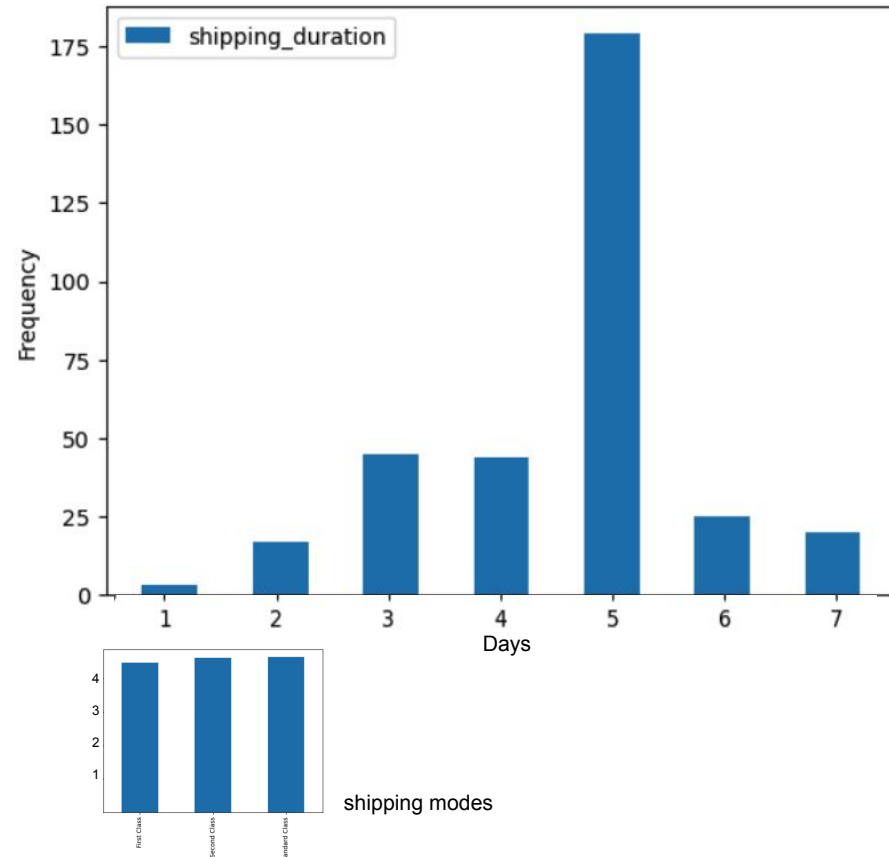
5 mode

1 minimum

7 maximum

Main observations

- ❑ Average duration was ~50% longer than expected
- ❑ 50% of the shippings needed 5 to 7 days
- ❑ There is no major difference in shipping modes



KPI 4 - Duration of overall process

Description

Overall time used for the complete process

Assumption [days]

n/a

Data frame

Muesli Project raw data.xlsx/ Campaign Data

Muesli Project raw data.xlsx/ Order Process Data

333 extracted samples

Main descriptions of data [days]

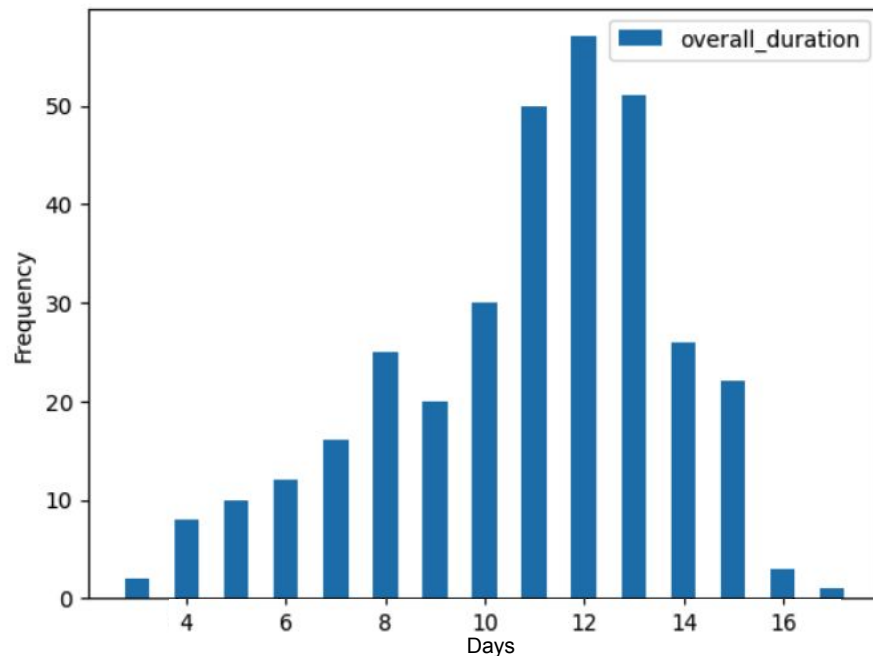
10.84 average

11 median

12 mode

3 minimum

17 maximum



Main observations

- ☐ High spread of durations

Further findings to consider

Description

Further findings to investigate more into:

- ❑ Shipping mode makes not a difference
- ❑ Outliner in the warehouse process do not follow any pattern, they are independent of day of order
- ❑ Profit of sales differs a lot, cost of production should be looked into
- ❑ Regarding the data we have, no difference in profit margin for express orders and non express orders
- ❑ Duration in warehouse - for product sub-categories

Main observations

- ❑ Huge difference between sub-categories

		process_time
category	sub-category	mean
Power Muesli	Hunger Buster	4.000000
	Maple Sweetened	2.000000
	Mega Protein	4.375000
	No Taste All Power	4.800000
	Nuts and more	3.931034
	Power Clean Fibre Boost	4.625000
	Steel Cut	4.684211
	Super Fibre Boost	3.967742
	Super Mega Protein	3.777778
	Gluten Free	4.352941
Special Projects Muesli	Low Sugar	2.333333
	Only Oats	3.416667
Toasted Muesli	Organic	5.500000
	Sweetened	3.833333
	With Fruit	5.142857
	With Fruit and Nuts	4.592593
	With Nuts	4.222222

Based on the findings we have the following first suggestions:

- ❑ Express mode is an option to improve customers service
- ❑ Consider to process orders received on Sunday automatically with express processing
- ❑ On Monday, the number of staff could be increased to process orders from the weekend faster or include weekend processing
- ❑ Outliers in the warehouse process do not follow a pattern, they are independent of the day of the order, the warehouse process should be monitored to draw better conclusions
- ❑ Take in consideration discussing with delivery company about their times



BACKUP

Duration in warehouse - product sub-categories

Description

Time between the order and the “ready to ship” time of different sub-categories

Data frame

Muesli Project raw data.xlsx/ InternDataStudy

Muesli Project raw data.xlsx/ Order Process Data

205 extracted samples

Main descriptions of data [days]

[refer to table](#)

		process_time
category	sub-category	mean
Power Muesli	Hunger Buster	4.000000
	Maple Sweetened	2.000000
	Mega Protein	4.375000
	No Taste All Power	4.800000
	Nuts and more	3.931034
	Power Clean Fibre Boost	4.625000
	Steel Cut	4.684211
	Super Fibre Boost	3.967742
	Super Mega Protein	3.777778
Special Projects Muesli	Gluten Free	4.352941
	Low Sugar	2.333333
	Only Oats	3.416667
	Organic	5.500000
Toasted Muesli	Sweetened	3.833333
	With Fruit	5.142857
	With Fruit and Nuts	4.592593
	With Nuts	4.222222

Main observations

- ❑ Huge difference between sub-categories