Delivery Time Report:

Muesli GmbH & Co. KG

A technical / code focused presentation
Louiza & Jamie
26.02.2024

Overview

Process Description

To fully understand the process from the moment of order until the delivery, we will have a look at a flowchart

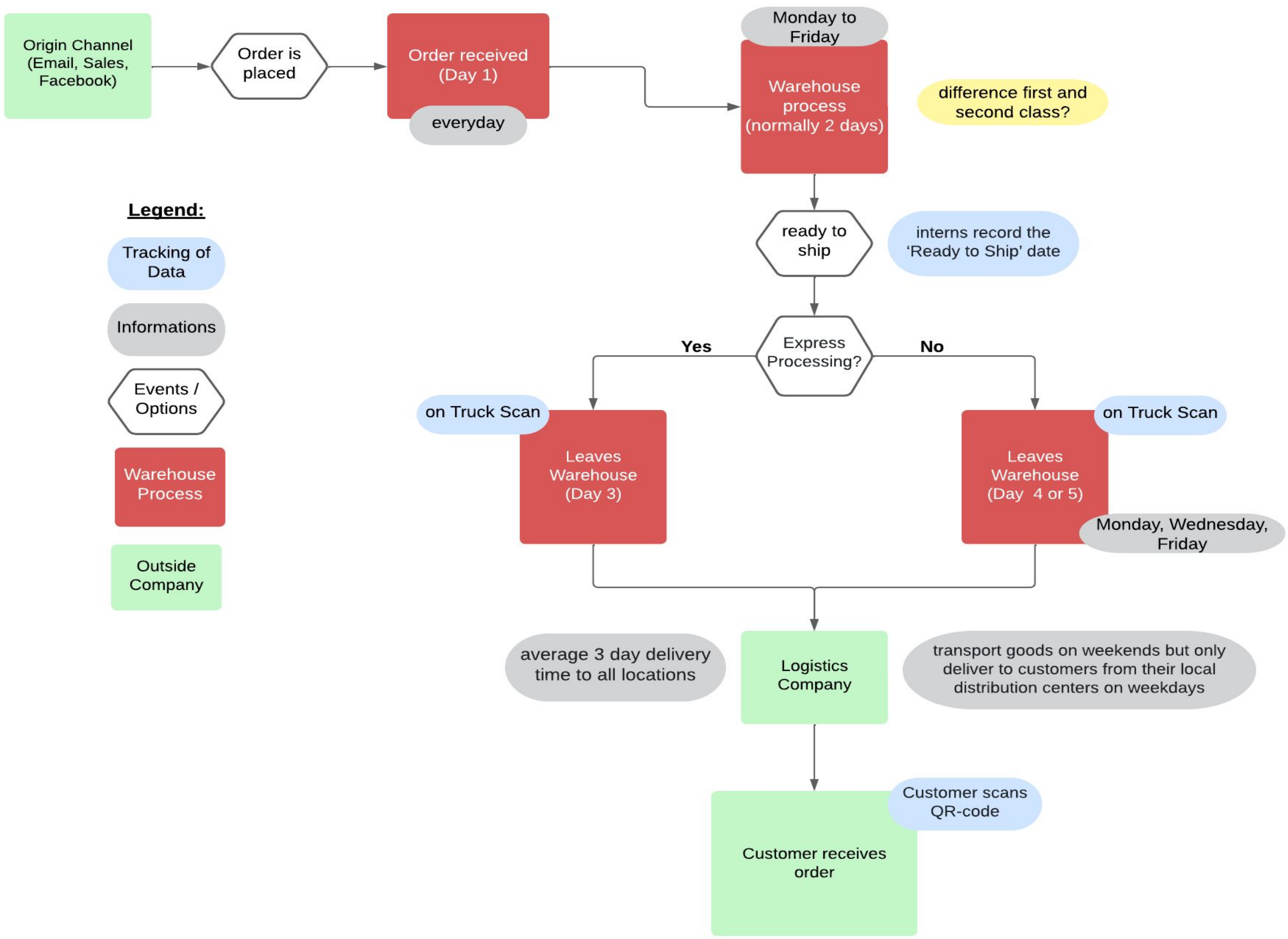
KPI’s

For a clear overview we are looking at 4 major KPI’s for the shipping and delivery process

Hypotheses

Our Data Analyst Team will face five hypotheses in order to get a greater understanding of the delivery process

Shipment Process



KPI's

- Order-lead Time
- Shipping Time
- Pack Cycle Time
- Pick Cycle

10 days 21 hrs

Average order lead time

order placed -> customer

```
order_lead = df_campaign.merge(df, how = 'left', on ='order_id')

order_lead['time'] = order_lead['arrival_scan_date'] - order_lead['order_date']
print(order_lead['time'].mean())
```

```
Output:
10 days 21:16:10.059880239
```

4 days 13 hrs

Average shipping time

(truck scan -> customer)

```
last_mile = df_campaign.merge(df_process, how = 'left', on ='order_id')
last_mile['time'] = last_mile ['arrival_scan_date'] - last_mile['on_truck_scan_date']

print(last_mile['time'].mean())

Output:
4 days 13:34:33.151750972
```

4 days 3 hrs

Average pack cycle time

order received -> ready to ship

```python
warehouse_pr['time'] = warehouse_pr['ready_to_ship_date'] -
warehouse_pr['order_date']

print(warehouse_pr['time'].mean())
```

```
Output:
4 days 03:47:22.105263157
```

1 day 15 hrs

Average pick cycle time

ready to ship -> truck scan

```
pick_cycle = df_intern.merge(df_process, how = 'left', on ='order_id')
pick_cycle['time'] = pick_cycle['pickup_date'] - pick_cycle['ready_to_ship_date']

print(pick_cycle['time'].mean())

Output:
1 days 15:26:41.913875598
```

Hypotheses

- Warehouse process might take longer than the two days as expected
- Last mile delivery time might be depending on delivery region
- The origin channel is influencing the process time
- The quantity of the order is influencing the process time
- The amount of orders per day is influencing the process time

1. Hypotheses:

The warehouse process might take longer than the 3 days expected

```
# create dataframe for internal warehouse process
warehouse_pr = df_intern.merge(df, how = 'left', on = 'order_id')
warehouse_pr = warehouse_pr[['order_id', 'origin_channel', 'quantity', 'ship_mode', 'order_date', 'ready_to_ship_date', 'ready_to_ship_day']]
warehouse_pr.drop_duplicates()

warehouse_pr['time'] = warehouse_pr['ready_to_ship_date'] - warehouse_pr['order_date']
print('the warehouse processing time is', warehouse_pr['time'].mean(),'\n')

# look at mean and count for ship_mode
print(warehouse_pr.groupby('ship_mode')["time"].mean(),'\n')
print(warehouse_pr.groupby('ship_mode')["time"].count(),'\n')

# create column for day of date
warehouse_pr['order_day'] = warehouse_pr['order_date'].dt.day_name()
df_intern["day"] = df_intern["ready_to_ship_date"].dt.day_name()
```

✓ 0.0s

```
the warehouse processing time is 4 days 03:47:22.105263157

ship_mode
First Class      1 days 21:52:56.470588235
Second Class     3 days 20:54:11.612903225
Standard Class   5 days 07:19:06.188340807
Name: time, dtype: timedelta64[ns]

ship_mode
First Class       102
Second Class       93
Standard Class    223
Name: time, dtype: int64
```

1. Hypotheses:

The warehouse process might take longer than the 3 days expected

Our hypothesis can be confirmed.

As we can see, our mean value for the internal processing time is roughly 4 days.

Furthermore the shipmode seems to have an influence on the processing time. We can see that there is deviation of roughly 2 days depending on the shipmode.

2. Hypotheses:

Last mile delivery time might be heavily depending on delivery region

```
last_mile = df_campaign.merge(df_process, how = 'left', on ='order_id')
last_mile['time'] = last_mile ['arrival_scan_date'] - last_mile['on_truck_scan_date']
last_mile

last_mile = last_mile.merge(df[['order_id','region']], how = 'left', on = 'order_id')

last_mile.groupby('region', as_index=False)['time'].mean()
```

✓ 0.0s

	region	time
0	Central	4 days 08:30:49.117174959
1	East	4 days 16:07:05.615763546
2	South	4 days 15:03:45.941422594
3	West	4 days 15:42:46.153846153

2. Hypotheses:

Last mile delivery time might be heavily depending on delivery region

we cannot see any notable dependency of the delivery region on the last mile delivery time. However we should mention that the process is, in average, roughly 7 hours faster for the central region

3. Hypotheses:

The origin channel is influencing the process time

```
origin_inf = pd.pivot_table(warehouse_pr,
                            values='time',
                            columns='origin_channel',
                            aggfunc='mean',
                            fill_value=0)
```

✓ 0.0s

```
origin_inf = origin_inf.T
origin_inf.reset_index(inplace = True)

origin_inf
```

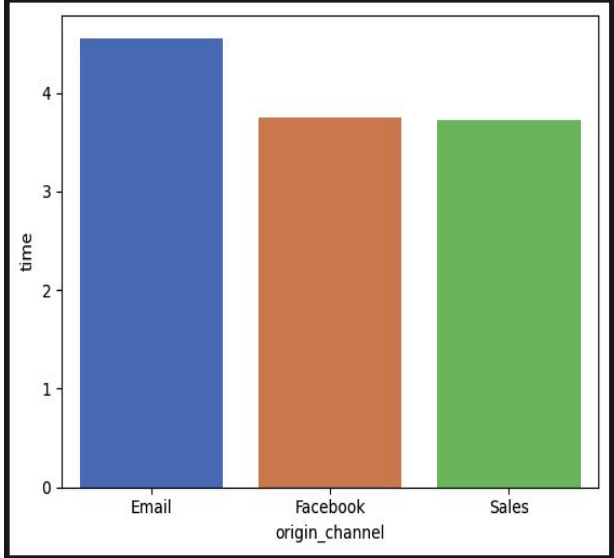
✓ 0.0s

	origin_channel	time
0	Email	4.558140
1	Facebook	3.752809
2	Sales	3.719298

```
sns.barplot(data = origin_inf, x = 'origin_channel', y = 'time', palette="muted")
```

✓ 0.1s

```
<Axes: xlabel='origin_channel', ylabel='time'>
```



3. Hypotheses:

The origin channel is influencing the process time

The order income via E-Mail tends to have a longer internal processing time of roughly one day compared to the other available channels.

4. Hypotheses:

The quantity of each order is influencing the process time

```
quantity_inf = pd.pivot_table(warehouse_pr,
                              values='time',
                              # index='time',
                              columns='quantity',
                              aggfunc='mean',
                              fill_value=0)
```

✓ 0.0s

```
quantity_inf = quantity_inf.T
quantity_inf.reset_index(inplace = True)

quantity_inf
```

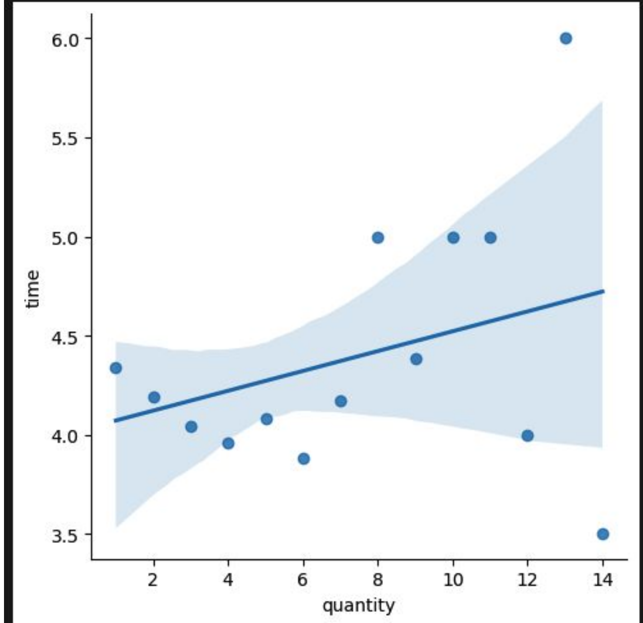
✓ 0.0s

	quantity	time
0	1	4.342105
1	2	4.190476
2	3	4.044444

```
sns.lmplot( data = quantity_inf, x = 'quantity', y = 'time', palette="muted")
```

✓ 0.1s

```
<seaborn.axisgrid.FacetGrid at 0x1135f5370>
```



4. Hypotheses:

The quantity of the order is influencing the process time

we can't examine any correlation between the quantity of products per order and internal processing time

5. Hypotheses:

The amount of orders per day is influencing the process time

```
df_time = pd.pivot_table( warehouse_pr,
                          values = 'time',
                          index='order_date',
                          aggfunc='mean',
                          fill_value=0)

counts = df['order_date'].value_counts()
df_numberorders = pd.DataFrame(counts.reset_index())

df_time_count = df_time.merge(df_numberorders, on = 'order_date')
df_time_count.sort_values('count', ascending= False)
```

✓ 0.0s

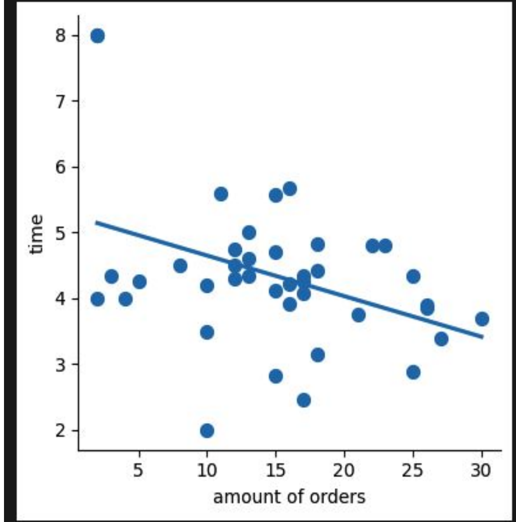
	order_date	time	count
23	2020-11-12	3.703704	30
17	2020-11-06	3.388889	27
0	2019-09-02	3.863636	26
30	2020-11-19	3.900000	26

```
fig=sns.lmplot(data = df_time_count, x = 'count', y = 'time', palette="muted", ci=None,
               height=4, scatter_kws={"s": 50, "alpha": 1})
fig.set_axis_labels('amount of orders', 'time')

print(df_time_count.corr())
```

✓ 0.1s

	order_date	time	count
order_date	1.000000	-0.038523	0.281849
time	-0.038523	1.000000	-0.390011
count	0.281849	-0.390011	1.000000



5. Hypotheses:

The amount of orders per day is influencing the process time

We were able to see that the warehouse process is slightly faster the more orders were coming in. However if we exclude the outlier the tilt of the graph becomes more horizontal.

Proposals

Trucks should leave everyday

Implement E-Mail automation

Increase warehouse efficiency

Thank you.