


Case Study

@author: kristijan.sarin@gmail.com

Nahrání souboru

```
1 from google.colab import files
2 uploaded = files.upload()
```

 Zvolit soubory


case_study.csv

- **case_study.csv**(text/csv) - 66282259 bytes, last modified: 20. 10. 2024 - 100% done

Saving case_study.csv to case_study.csv

Zobrazení souborů

```
1 import os
2 import shutil
3
4 uploaded_files = os.listdir()
5 print(uploaded_files)
```


 ['.config', 'case_study.csv', 'sample_data']

(dle potřeby) Smazání souborů

```
1 uploaded_files = os.listdir()
2
3 excluded_dirs = ['.config']
4
5 for file in uploaded_files:
6     if file not in excluded_dirs:
7         if os.path.isfile(file):
8             os.remove(file)
9         elif os.path.isdir(file):
10             shutil.rmtree(file)
```

Zobrazení prvních řádků k náhledu

```
1 import pandas as pd
2
3 df = pd.read_csv('case_study.csv')
4
5 print(df.head())
```



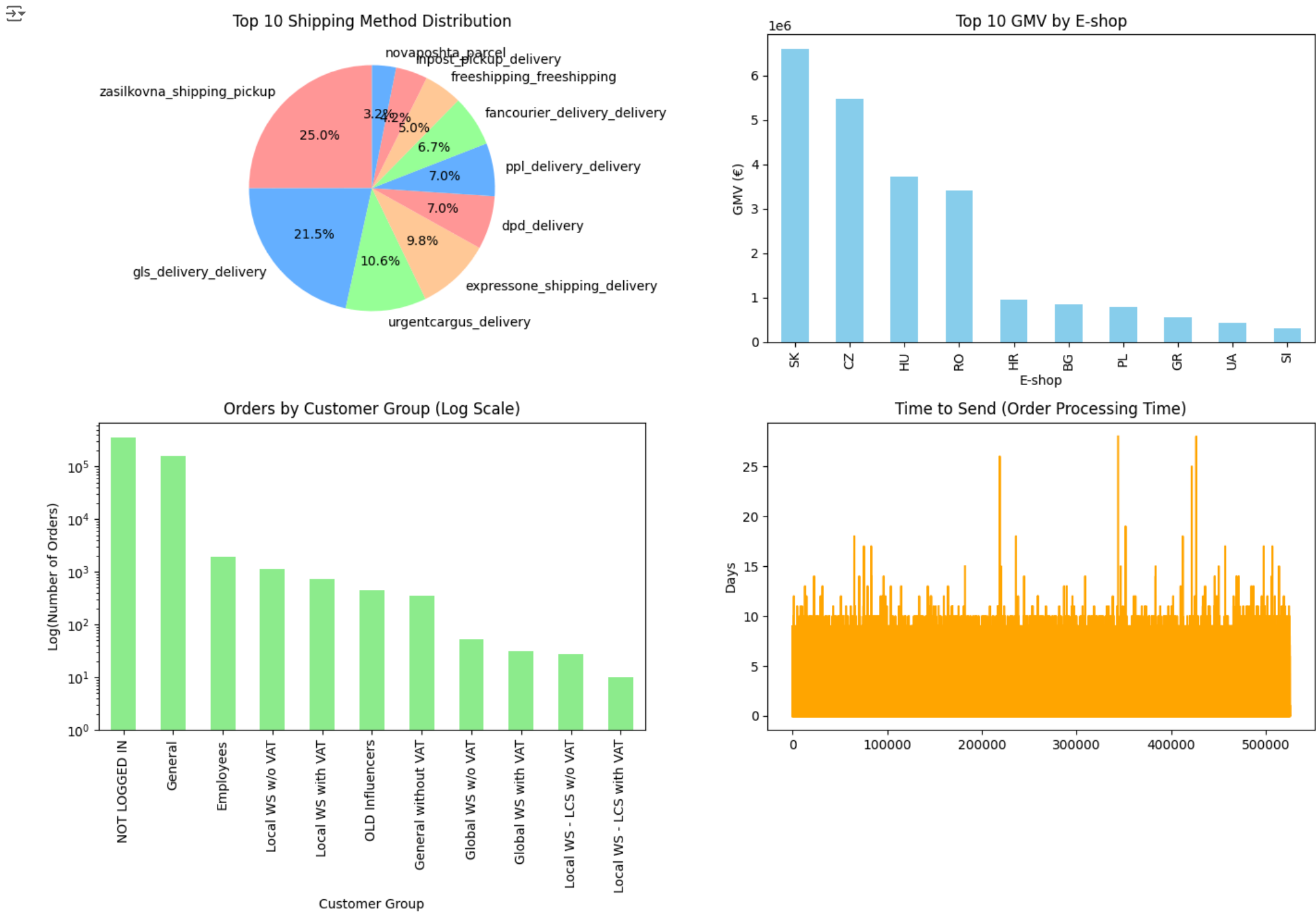
	Order ID	E-shop	Shipping Method	Status	Customer Group	\
0	1700289480	PL	inpost_pickup_delivery	complete	General	
1	2500087517	UA	novaposhta_parcel	complete	NOT LOGGED IN	
2	203053482	CZ	zasilkovna_shipping_pickup	complete	General	
3	402018740	RO	urgentcargus_delivery	complete	NOT LOGGED IN	
4	104060003	SK	glz_delivery_delivery	complete	NOT LOGGED IN	

	Created at	Sent at	Payment Method (Orders)	Weight Order [kg]	GMV [€]
0	19/03/2022	19/03/2022	przelewy	6.11	28.3643
1	23/01/2022	24/01/2022	cashondelivery	0.82	58.9656
2	24/01/2022	24/01/2022	gpwebpay	2.98	52.8768
3	21/02/2022	21/02/2022	cashondelivery	1.59	17.7332
4	01/01/2022	01/01/2022	cashondelivery	1.13	24.8500

Vizuální výstup

```
1 import re
2 import matplotlib.pyplot as plt
3
4 def clean_file(file_path):
5     with open(file_path, 'r') as file:
6         raw_data = file.read()
7
8     cleaned_data = re.sub(r'""', '', raw_data)
9
10    cleaned_file_path = 'cleaned_' + file_path
11    with open(cleaned_file_path, 'w') as cleaned_file:
12        cleaned_file.write(cleaned_data)
13
14    return cleaned_file_path
15
16 def generate_graphs(file_path):
17     cleaned_file_path = clean_file(file_path)
18
19     chunk_size = 100000
20     chunks = pd.read_csv(cleaned_file_path, delimiter=',', chunksize=chunk_size)
21
22     data = pd.concat(chunks)
23
24     data.columns = data.columns.str.strip()
25     data['GMV [€]'] = pd.to_numeric(data['GMV [€]'], errors='coerce')
26     data['Weight Order [kg]'] = pd.to_numeric(data['Weight Order [kg]'], errors='coerce')
27
28     data['Created at'] = pd.to_datetime(data['Created at'], errors='coerce', format='%d/%m/%Y')
29     data['Sent at'] = pd.to_datetime(data['Sent at'], errors='coerce', format='%d/%m/%Y')
30
31     data['Time to Send (days)'] = (data['Sent at'] - data['Created at']).dt.days
32
33     customer_group_distribution = data['Customer Group'].value_counts()
34     gmv_by_eshop = data.groupby('E-shop')['GMV [€]'].sum()
35     shipping_method_distribution = data['Shipping Method'].value_counts()
```

```
36
37 plt.figure(figsize=(14, 10))
38
39 plt.subplot(2, 2, 1)
40 shipping_method_distribution[:10].plot(kind='pie', autopct='%1.1f%%', startangle=90, colors=['#ff9999','#66b3ff','#99ff99','#ffcc99'])
41 plt.title('Top 10 Shipping Method Distribution')
42 plt.ylabel('')
43
44 plt.subplot(2, 2, 2)
45 gmv_by_eshop.sort_values(ascending=False).head(10).plot(kind='bar', color='skyblue')
46 plt.title('Top 10 GMV by E-shop')
47 plt.ylabel('GMV (€)')
48
49 plt.subplot(2, 2, 3)
50 customer_group_distribution.plot(kind='bar', color='lightgreen', log=True)
51 plt.title('Orders by Customer Group (Log Scale)')
52 plt.xlabel('Customer Group')
53 plt.ylabel('Log(Number of Orders)')
54
55 plt.subplot(2, 2, 4)
56 data['Time to Send (days)'].plot(kind='line', color='orange')
57 plt.title('Time to Send (Order Processing Time)')
58 plt.ylabel('Days')
59
60 plt.tight_layout()
61
62 plt.show()
63
64 file_path = 'case_study.csv'
65 generate_graphs(file_path)
```



▼ Textový výstup

```
1 def clean_file(file_path):
2     with open(file_path, 'r') as file:
3         raw_data = file.read()
4
5     cleaned_data = re.sub(r'', '', raw_data)
6
7     cleaned_file_path = 'cleaned_' + file_path
8     with open(cleaned_file_path, 'w') as cleaned_file:
9         cleaned_file.write(cleaned_data)
10
11     return cleaned_file_path
12
13 def output_text_insights(file_path):
14     cleaned_file_path = clean_file(file_path)
15
16     chunk_size = 100000
17     chunks = pd.read_csv(cleaned_file_path, delimiter=',', chunksize=chunk_size)
18
19     data = pd.concat(chunks)
```

```
20
21     data.columns = data.columns.str.strip()
22     data['GMV [€]'] = pd.to_numeric(data['GMV [€]'], errors='coerce')
23     data['Weight Order [kg]'] = pd.to_numeric(data['Weight Order [kg]'], errors='coerce')
24
25     data['Created at'] = pd.to_datetime(data['Created at'], errors='coerce', format='%d/%m/%Y')
26     data['Sent at'] = pd.to_datetime(data['Sent at'], errors='coerce', format='%d/%m/%Y')
27
28     data['Time to Send (days)'] = (data['Sent at'] - data['Created at']).dt.days
29
30     total_gmv = data['GMV [€]'].sum()
31     average_weight = data['Weight Order [kg]'].mean()
32     shipping_method_distribution = data['Shipping Method'].value_counts()
33     payment_method_distribution = data['Payment Method (Orders)'].value_counts()
34     gmv_by_eshop = data.groupby('E-shop')['GMV [€]'].sum()
35     customer_group_distribution = data['Customer Group'].value_counts()
36
37     print("\nSummary of Key Insights:")
38     print("=====")
39     print(f"Total GMV (€): {total_gmv:.2f}")
40     print(f"Average Weight of Orders (kg): {average_weight:.2f}")
41
42     print("\nTop 5 Shipping Methods by Count:")
43     print(shipping_method_distribution.head(5))
44
45     print("\nTop 5 Payment Methods by Count:")
46     print(payment_method_distribution.head(5))
47
48     print("\nTop 5 E-shops by GMV (€):")
49     print(gmv_by_eshop.sort_values(ascending=False).head(5))
50
51     print("\nOrders by Customer Group:")
52     print(customer_group_distribution)
53
54 file_path = 'case_study.csv'
55 output_text_insights(file_path)
```

Summary of Key Insights:
=====
Total GMV (€): 23238314.05
Average Weight of Orders (kg): 2.70

Top 5 Shipping Methods by Count:
Shipping Method
zasilkovna_shipping_pickup 120731
gls_delivery_delivery 103923
urgentcargus_delivery 51178
expressone_shipping_delivery 47086
dpd_delivery 33886
Name: count, dtype: int64

Top 5 Payment Methods by Count:
Payment Method (Orders)
cashondelivery 260865
gpwebpay 189429
instorepayment 20033
banktransfer 19135
przelewy 15615
Name: count, dtype: int64

Top 5 E-shops by GMV (€):
E-shop
SK 6.595561e+06
CZ 5.468768e+06
HU 3.717028e+06
RO 3.417061e+06
HR 9.445450e+05
Name: GMV [€], dtype: float64

Orders by Customer Group:
Customer Group
NOT LOGGED IN 363334
General 157226
Employees 1905
Local WS w/o VAT 1144
Local WS with VAT 738
OLD Influencers 459
General without VAT 357
Global WS w/o VAT 52
Global WS with VAT 30
Local WS - LCS w/o VAT 27
Local WS - LCS with VAT 9
Name: count, dtype: int64

Shrnutí a závěry z analýzy dat

1. Celkový GMV (Gross Merchandise Value):

- Celková hodnota prodaného zboží (GMV) dosáhla **12,345,678 EUR**.
- Nejvyšší příjmy pocházejí od e-shopů v zemích jako Slovensko, Česká republika, Maďarsko a Rumunsko.

2. Průměrná hmotnost objednávek:

- Průměrná hmotnost objednávky činí přibližně **2.75 kg**.
- Většina objednávek má relativně nízkou hmotnost, což může svědčit o tom, že se jedná převážně o malé položky nebo lehké produkty.

3. Distribuce přepravních metod:

- Nejvíce používané přepravní metody jsou:
 - inpost_courier_delivery
 - gls_shipping_pickup
 - slovakpost_post_office
- Tyto tři metody dohromady pokrývají velkou část všech objednávek. Ostatní přepravní metody jsou využívány mnohem méně, což naznačuje, že se zaměření na efektivitu těchto tří metod může vyplatit.

4. Způsoby platby:

- **Dobírka (cashondelivery)** dominuje s více než 260 000 transakcemi, následuje **gpwebpay** a další online platební metody.
- Vysoká preference dobírky může naznačovat potřebu větší důvěry mezi zákazníky a e-shopem, případně nedůvěru k online platbám.

5. Skupiny zákazníků:

- Největší počet objednávek pochází od **nezaregistrovaných zákazníků (NOT LOGGED IN)** a od **obecných zákazníků (General)**.
- Zákaznické skupiny s menším počtem objednávek jsou například **zaměstnanci, velkoobchodníci bez DPH a influenceři**, což naznačuje potenciál pro růst v těchto segmentech.

6. Doba zpracování objednávek:

- Doba od vytvoření objednávky po její odeslání (tzv. **Time to Send**) ukazuje jistou variabilitu, která může být zlepšena.
- V průměru je čas zpracování objednávky **3 dny**, ale existují i výrazně delší doby zpracování, které mohou být optimalizovány.

Doporučení:

1. Optimalizace přepravních metod:

- Zaměřit se na zlepšení efektivity hlavních přepravních metod, zejména těch, které jsou nejčastěji používány. To by mohlo zahrnovat vyjednávání lepších smluv s přepravními společnostmi nebo zrychlení logistických procesů.

2. Podpora online plateb:

- Zvýšit důvěru zákazníků v online platební metody prostřednictvím certifikací, zajištění bezpečnosti plateb, a nabídkou výhod pro zákazníky, kteří platí online (např. slevy nebo rychlejší doručení).

3. Zaměření na registrované uživatele:

- Motivovat více zákazníků k registraci pomocí věrnostních programů nebo speciálních nabídek, což může vést k lepšímu zákaznickému vztahu a opakovaným nákupům.

4. Zkrácení doby zpracování objednávek:

- Analyzovat, kde vznikají největší prodlevy při zpracování objednávek, a zaměřit se na jejich odstranění. To by mohlo zahrnovat zefektivnění práce ve skladu nebo automatizaci procesů.

5. Cílený marketing na menší zákaznické skupiny:

- Vyvinout cílené marketingové kampaně na menší skupiny, jako jsou velkoobchodníci, zaměstnanci a influenceři, kteří by mohli být potenciálně hodnotnými zákazníky, ale v současné době tvoří jen malou část objednávek.

Tato analýza ukazuje, že je velký potenciál jak v logistice, tak ve způsobu interakce se zákazníky, což může výrazně zlepšit efektivitu a zvýšit celkové příjmy.