Prepare data

```
from google.colab import files

uploaded = files.upload()
```

Choose Files Kasatria Ent…Test v2.xlsx

- **Kasatria Entrance Test v2.xlsx**(application/vnd.openxmlformats-officedocument.spreadsheetml.sheet) - 645738 bytes, last modified: 9/22/2022 - 100% done

Saving Kasatria Entrance Test v2.xlsx to Kasatria Entrance Test v2.xlsx

```
# Import libraries
import pandas as pd
import io
import datetime as dt
```

```
# Import Orders sheet
orders = pd.read_excel(io.BytesIO(uploaded['Kasatria Entrance Test v2.xlsx']), sheet_name='Orders')
```

```
# Review Orders
orders.head(6)
```

	Order ID	Order Priority	Order Date	Ship Date	Item Type	Units Sold	Price	Cost	Sales Channel	Country
0	506209075	H	2010-01-01	2010-02-04	Vegetables	7369	154.06	90.93	Online	Central African Republic
1	863776719	C	2010-01-01	2010-02-10	Cereal	9581	205.70	117.11	Online	China
2	695167052	C	2010-01-02	2010-01-22	Cosmetics	4234	437.20	263.33	Offline	Mongolia
			2010-	2010-						Equatorial

```
orders.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10000 entries, 0 to 9999
Data columns (total 10 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   Order ID        10000 non-null  int64
 1   Order Priority  10000 non-null  object
 2   Order Date      10000 non-null  datetime64[ns]
 3   Ship Date       10000 non-null  datetime64[ns]
 4   Item Type       10000 non-null  object
 5   Units Sold      10000 non-null  int64
 6   Price           10000 non-null  float64
 7   Cost            10000 non-null  float64
 8   Sales Channel   10000 non-null  object
 9   Country         10000 non-null  object
dtypes: datetime64[ns](2), float64(2), int64(2), object(4)
memory usage: 781.4+ KB
```

```
orders['Order ID'].nunique()
```

```
10000
```

```
orders.describe(include='object')
```

	Order Priority	Item Type	Sales Channel	Country
count	10000	10000	10000	10000
unique	4	12	2	185
top	C	Personal Care	Online	United Kingdom
freq	2555	888	5061	72

```
# Remove spaces from Country column
orders['Country'] = orders['Country'].str.strip()
```

```
# Import Regions sheet
regions = pd.read_excel(io.BytesIO(uploaded['Kasatria Entrance Test v2.xlsx']), sheet_name='Regions'
```

```
# Review Regions
regions
```

	Region	Countries
0	Asia	China; Mongolia; Uzbekistan; Laos; Maldives; T...
1	Australia and Oceania	Tonga; Kiribati; Solomon Islands; Palau; New Z...
2	Central America and the Caribbean	Haiti; Dominica; Guatemala; The Bahamas; Grena...
3	Europe	Sweden; Kosovo; Iceland; France; Latvia; Russi...
4	Middle East and North Africa	Oman; Morocco; Iraq; Egypt; Algeria; Saudi Ara...
5	North America	United States of America; Canada; Greenland; M...
6	Sub-Saharan Africa	Central African Republic; Equatorial Guinea; S...

```
# Split all countries in Countries column
countries = regions['Countries'].str.split(';', expand=True)
# Join countries and regions
region_countries = regions[['Region']].merge(countries, left_index=True, right_index=True)
region_countries
```

	Region	0	1	2	3	4	5	6	
0	Asia	China	Mongolia	Uzbekistan	Laos	Maldives	Taiwan	Kazakhstan	Sri
1	Australia and Oceania	Tonga	Kiribati	Solomon Islands	Palau	New Zealand	Federated States of Micronesia	East Timor	
2	Central America and the Caribbean	Haiti	Dominica	Guatemala	The Bahamas	Grenada	Antigua and Barbuda	Costa Rica	Dor R
3	Europe	Sweden	Kosovo	Iceland	France	Latvia	Russia	San Marino	Liecht
4	Middle East and North Africa	Oman	Morocco	Iraq	Egypt	Algeria	Saudi Arabia	Yemen	P
	North	United States							

```
# Reshape region_countries from wide to long
regions_melt = pd.melt(region_countries, id_vars='Region', value_name='Country')
regions_melt.head()
```

	Region	variable	Country
0	Asia	0	China
1	Australia and Oceania	0	Tonga
2	Central America and the Caribbean	0	Haiti
3	Europe	0	Sweden
4	Middle East and North Africa	0	Oman

```
# Keep 2 columns: Country includes all unique countries with their corresponding region, drop null v
regions_melt = regions_melt[['Region', 'Country']].dropna(subset=['Country'])
# Remove spaces from Country column
regions_melt['Country'] = regions_melt['Country'].str.strip()
regions_melt.head()
```

	Region	Country
0	Asia	China
1	Australia and Oceania	Tonga
2	Central America and the Caribbean	Haiti
3	Europe	Sweden
4	Middle East and North Africa	Oman

```
# Check regions_melt table
regions_melt.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
Int64Index: 185 entries, 0 to 335
Data columns (total 2 columns):
 #   Column   Non-Null Count  Dtype 
---  ------   --------------  ----- 
 0   Region   185 non-null    object
 1   Country  185 non-null    object
dtypes: object(2)
memory usage: 4.3+ KB
```

```
# Create data table by joining 2 tables: orders and regions_melt on Country column
data = orders.merge(regions_melt, on='Country', how='inner')
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 10000 entries, 0 to 9999
Data columns (total 11 columns):
 #   Column          Non-Null Count  Dtype         
---  ------          --------------  -----         
 0   Order ID        10000 non-null  int64         
 1   Order Priority  10000 non-null  object        
 2   Order Date      10000 non-null  datetime64[ns]
 3   Ship Date       10000 non-null  datetime64[ns]
 4   Item Type       10000 non-null  object        
 5   Units Sold      10000 non-null  int64         
 6   Price           10000 non-null  float64       
 7   Cost            10000 non-null  float64       
 8   Sales Channel   10000 non-null  object        
 9   Country         10000 non-null  object        
 10  Region          10000 non-null  object        
dtypes: datetime64[ns](2), float64(2), int64(2), object(5)
memory usage: 937.5+ KB
```

Answer questions

1. Total cost, revenue and profit by both region and item type?

```
data['total_cost'] = data['Units Sold'] * data['Cost']
data['total_price'] = data['Units Sold'] * data['Price']
data['profit'] = data['total_price'] - data['total_cost']
data.groupby(['Region','Item Type'])[['total_cost', 'total_price', 'profit']].sum()
```

1 to 84 of 84 entries Filter

Region	Item Type	total_cost	total_price	profit
Asia	Baby Food	106089386.82	169881436.88	63792050.06000001
Asia	Beverages	17553897.57	26201083.35	8647185.780000003
Asia	Cereal	70979199.89999999	124672713.0	53693513.099999994
Asia	Clothes	18966312.96	57830320.32	38864007.36
Asia	Cosmetics	156914923.70999998	260521796.39999998	103606872.69
Asia	Fruits	3595133.76	4847196.24	1252062.48
Asia	Household	296788063.04	394664223.52	97876160.47999997
Asia	Meat	207531950.16	240082410.95999998	32550460.799999993
Asia	Office Supplies	374430869.76000005	464479439.76000005	90048570.0
Asia	Personal Care	38511741.93	55541991.67	17030249.740000002
Asia	Snacks	60843874.559999995	95274613.92	34430739.36000001
Asia	Vegetables	65696925.00000001	111308350.0	45611425.0
Australia and Oceania	Baby Food	53987423.58	86450316.72	32462893.140000008
Australia and Oceania	Beverages	7368159.04	10997771.200000001	3629612.160000001
Australia and Oceania	Cereal	44082780.42	77430005.39999999	33347224.979999997
Australia and Oceania	Clothes	14071715.840000002	42906169.28	28834453.439999998
Australia and Oceania	Cosmetics	102387970.6	169992104.0	67604133.4
Australia and Oceania	Fruits	2143414.64	2889892.86	746478.22
Australia and Oceania	Household	185950353.34	247273933.67	61323580.32999998
Australia and Oceania	Meat	109901519.64	127139082.83999999	17237563.199999996
Australia and Oceania	Office Supplies	121274159.36000001	150439929.36	29165770.0
Australia and Oceania	Personal Care	19217987.07	27716359.330000002	8498372.26
Australia and Oceania	Snacks	32296488.0	50572641.00000001	18276153.000000007
Australia and Oceania	Vegetables	33049054.080000002	55994031.36	22944977.279999997
Central America and the Caribbean	Baby Food	54587799.3	87411701.2	32823901.900000006
Central America and the Caribbean	Beverages	14615071.02	21814568.1	7199497.080000002
Central America and the Caribbean	Cereal	50663542.65	88988905.5	38325362.849999994
Central America and the Caribbean	Clothes	15280419.840000002	46591637.28	31311217.439999998
Central America and the Caribbean	Cosmetics	85972768.39	142738367.6	56765599.21
Central America and the Caribbean	Fruits	2864084.2	3861547.05	997462.8500000001
Central America and the Caribbean	Household	269079515.06	357817820.53	88738305.46999998
Central America and the Caribbean	Meat	149072872.54	172454287.74	23381415.2
Central America and the Caribbean	Office Supplies	236120708.48000002	292906443.48	56785735.0
Central America and the Caribbean	Personal Care	24515385.330000002	35356316.27	10840930.940000001
Central America and the Caribbean	Snacks	45793487.04	71707412.28	25913925.240000006
Central America and the Caribbean	Vegetables	43606208.940000005	73880705.48	30274496.54
Europe	Baby Food	161105388.23999998	257978820.16	96873431.92000002
Europe	Beverages	29899448.7	44628148.5	14728699.800000003
Europe	Cereal	133637148.75	234729412.5	101092263.74999999
Europe	Clothes	40648867.84	123942753.28	83293885.44
Europe	Cosmetics	301384344.96	500380646.4	198996301.44
Europe	Fruits	7283694.4399999995	9820356.81	2536662.37
Europe	Household	516393017.64000005	686691530.8199999	170298513.17999995
Europe	Meat	409975745.44	474278612.64	64302867.19999999
Europe	Office Supplies	591837804.1600001	734171549.1600001	142333745.0
Europe	Personal Care	66489664.260000005	95892010.94	29402346.680000003
Europe	Snacks	99232603.67999999	155387014.26000002	56154410.58000001
Europe	Vegetables	96484731.84	163471217.28	66986485.44
Middle East and North Africa	Baby Food	84841410.96	135856952.64000002	51015541.68000001
Middle East and North Africa	Beverages	16311417.209999999	24346547.55	8035130.340000002
Middle East and North Africa	Cereal	59891107.99	105196831.3	45305723.309999995
Middle East and North Africa	Clothes	21748106.240000002	66312306.08	44564199.839999996
Middle East and North Africa	Cosmetics	133982830.66	222448234.4	88465403.74000001
Middle East and North Africa	Fruits	3633567.44	4899015.06	1265447.62
Middle East and North Africa	Household	310916472.6	413451966.3	102535493.69999999
Middle East and North Africa	Meat	177596006.82	205451148.42	27855141.599999994
Middle East and North Africa	Office Supplies	263832821.92000002	327283168.17	63450346.25
Middle East and North Africa	Personal Care	35127692.88	50661484.72	15533791.840000002
Middle East and North Africa	Snacks	46902549.12	73444077.84	26541528.720000006
Middle East and North Africa	Vegetables	50925619.29000001	86281765.18	35356145.89
North America	Baby Food	20596904.58	32981920.72	12385016.140000002
North America	Beverages	2992519.86	4466658.3	1474138.4400000004

North America	**Cereal**	8399012.09	14752598.299999999	6353586.209999999
North America	**Clothes**	3920035.8400000003	11952609.28	8032573.44
North America	**Cosmetics**	32397753.229999997	53789153.199999996	21391399.970000003
North America	**Fruits**	452796.36	610489.89	157693.53000000003
North America	**Household**	65157828.78	86645883.39	21488054.609999992
North America	**Meat**	34599963.75	40026813.75	5426850.0
North America	**Office Supplies**	47549826.88	58985299.38	11435472.5
North America	**Personal Care**	5351518.109999999	7718009.09	2366490.9800000004
North America	**Snacks**	8102038.56	12686874.420000002	4584835.860000001
North America	**Vegetables**	6336729.840000001	10736133.28	4399403.4399999995
Sub-Saharan Africa	**Baby Food**	187878383.04	300850543.36	112972160.32000002
Sub-Saharan Africa	**Beverages**	35572247.04	53095411.2	17523164.160000004
Sub-Saharan Africa	**Cereal**	124606679.54	218867679.79999998	94261000.25999999
Sub-Saharan Africa	**Clothes**	41352765.440000005	126089012.48	84736247.03999999
Sub-Saharan Africa	**Cosmetics**	267478764.14999998	444088086.0	176609321.85
Sub-Saharan Africa	**Fruits**	8094282.4799999995	10913245.02	2818962.54
Sub-Saharan Africa	**Household**	535131729.16	711609982.5799999	176478253.41999996
Sub-Saharan Africa	**Meat**	364753820.75	421963830.75	57210009.999999985
Sub-Saharan Africa	**Office Supplies**	527927073.92	654890638.9200001	126963565.0
Sub-Saharan Africa	**Personal Care**	60294216.51	86956878.69	26662662.180000003
Sub-Saharan Africa	**Snacks**	108552739.67999999	169981291.26000002	61428551.58000001
Sub-Saharan Africa	**Vegetables**	89270982.15	151249175.3	61978193.15

2. How many orders of Beverages are there in 2011?

```
data['Order_Year'] = data['Order Date'].dt.to_period('Y')
data[(data['Item Type'] == 'Beverages') & (data['Order_Year'] == '2011')]['Order ID'].count()
```

```
109
```

It's 109 orders

3. For each item type, what is the country which gains the max profit?

```
# Create item_country table shows total profit by Item Type and Country
item_country = pd.DataFrame(data.groupby(['Item Type', 'Country'])['profit'].sum())
item_country = item_country.reset_index().sort_values(by=['Item Type', 'profit'], ascending=[True,Fa
# Rank profit
item_country['rank'] = item_country.groupby(['Item Type'])['profit'].rank(method='first', ascending=
item_country[item_country['rank'] == 1]
```

	Item Type	Country	profit	rank
148	Baby Food	Somalia	6044356.44	1.0
343	Beverages	Taiwan	910237.50	1.0
379	Cereal	Bangladesh	6008793.93	1.0
573	Clothes	Burundi	4347354.24	1.0
864	Cosmetics	Qatar	14081731.30	1.0
1099	Fruits	Zimbabwe	156295.73	1.0
1217	Household	Nigeria	10172341.67	1.0
1437	Meat	Sudan	3231971.60	1.0
1556	Office Supplies	Liechtenstein	7569950.00	1.0
1765	Personal Care	Nicaragua	1431828.16	1.0
1937	Snacks	Moldova	3599042.94	1.0
2050	Vegetables	Croatia	3947771.42	1.0

4. Which region has longest average delivery time in 2016? How long?

```
data['delivery_time'] = (data['Ship Date'] - data['Order Date']).astype('timedelta64[h]') / 24
data[data['Order_Year'] == '2016'].groupby('Region')['delivery_time'].mean().sort_values(ascending=F
```

```
Region
North America    27.857143
Name: delivery_time, dtype: float64
```

```
It's North America, 27.857143 days in average
```

5. Which item type contributes most profit in Jan?

```
data['order_month'] = data['Order Date'].dt.strftime('%m')
data['order_month']
data[data['order_month'] == '01'].groupby('Item Type')['profit'].sum().sort_values(ascending=False)
```

```
Item Type
```

```
Household          60246998.25
Cosmetics          58560111.48
Office Supplies    43323192.50
Baby Food          33403375.60
Cereal             32768023.56
Vegetables         25657925.90
Snacks             24649454.76
Clothes            22830219.36
Meat               21572579.60
Personal Care      10668668.50
Beverages           4599232.38
Fruits               795545.82
Name: profit, dtype: float64
```

```
data[data['order_month'] == '01'].groupby('Item Type')['profit'].sum().sort_values(ascending=False).
```

```
Item Type
Household    17.768028
Name: profit, dtype: float64
```

```
It's Household, which accounts for 17.77%
```

6. What is total profit of top 5 countries which are sorted by Online channel orders?

```
# top_5_online_orders table shows top 5 countries which are sorted by Online channel orders
top_5_online_orders = data[data['Sales Channel'] == 'Online'].groupby('Country')['Order ID'].count()
top_5_online_orders = pd.DataFrame(top_5_online_orders).reset_index()
# total_profit table shows total profit by countries
total_profit = data.groupby('Country')['profit'].sum()
total_profit = pd.DataFrame(total_profit).reset_index()
# Join top_5_online_orders and total_profit, sort values by number of Online Order ID
total_profit.merge(top_5_online_orders, how='inner', on='Country').sort_values(by='Order ID', ascend
```

	Country	profit	Order ID
3	Lithuania	28063374.02	42
0	Cambodia	23744162.89	41
1	Croatia	26096999.84	39
4	United Kingdom	25528037.97	39
2	India	24289892.46	38

7. Evaluate business situation. Illustrate your insights by chart. In your opinion, what does the company should improvise?

```
# Add profit_to_revenue column, ratio of profit to revenue earned for each order.
data['profit_to_revenue'] = data['profit'] / data['total_price']
```

To answer this question, I use Power BI to visualize data from "data" table and look at the following metrics to evaluate the business:

1. Delivery time
2. Profit
3. Profit_to_revenue
4. Profit growth: showing the change of profit year to year.

Calculate profit growth by Power BI measure:

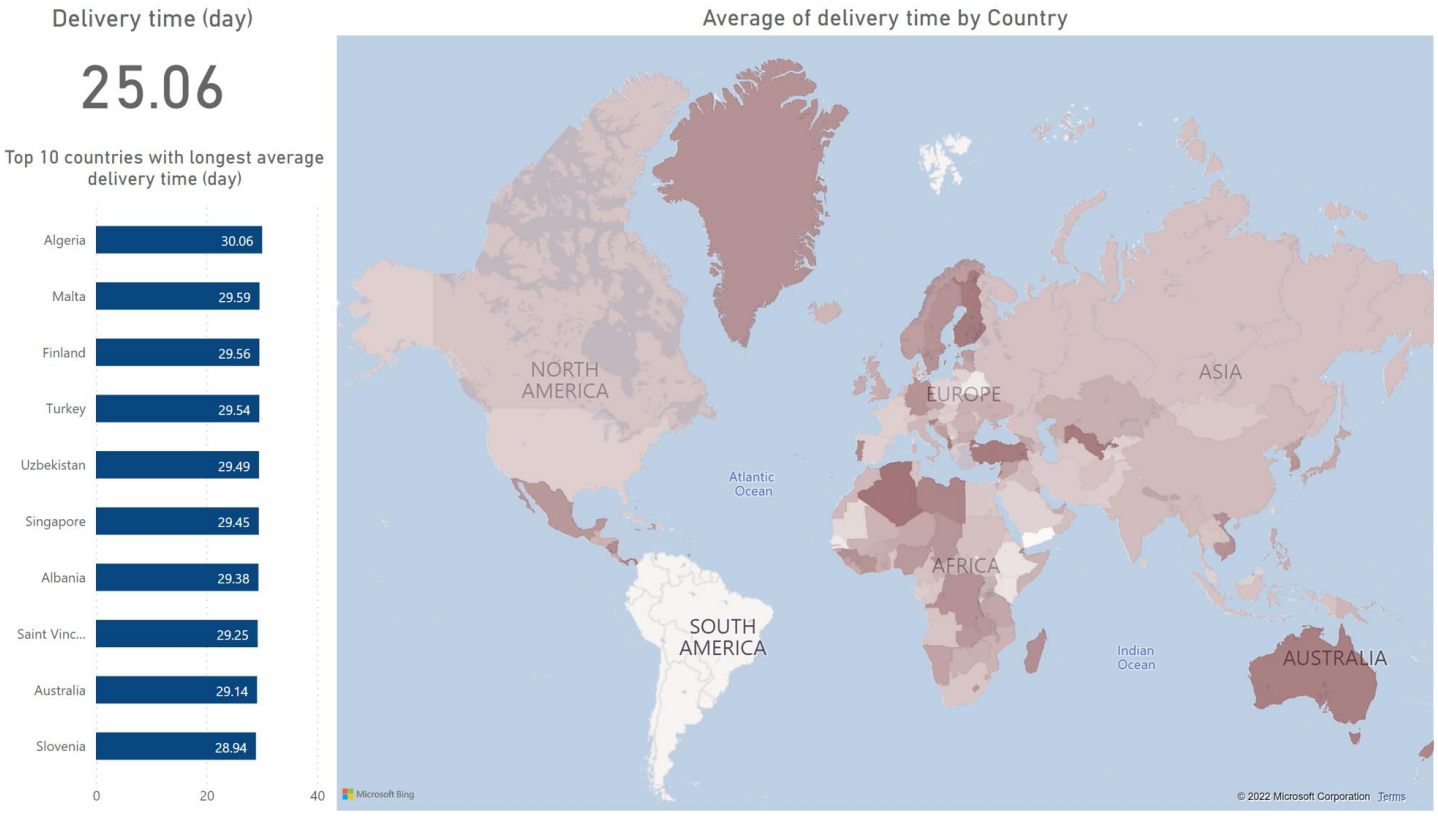
```
profit growth = IF( ISFILTERED('data'[Order Date]), ERROR("Time intelligence quick measures can only be grouped or filtered by the Power BI-provided date hierarchy or primary date column."), VAR __PREV_YEAR =

        CALCULATE(
          SUM('data'[profit]),
          DATEADD('data'[Order Date].[Date], -1, YEAR))

      RETURN

       DIVIDE(SUM('data'[profit]) - __PREV_YEAR, __PREV_YEAR))
```

1. Delivery time:

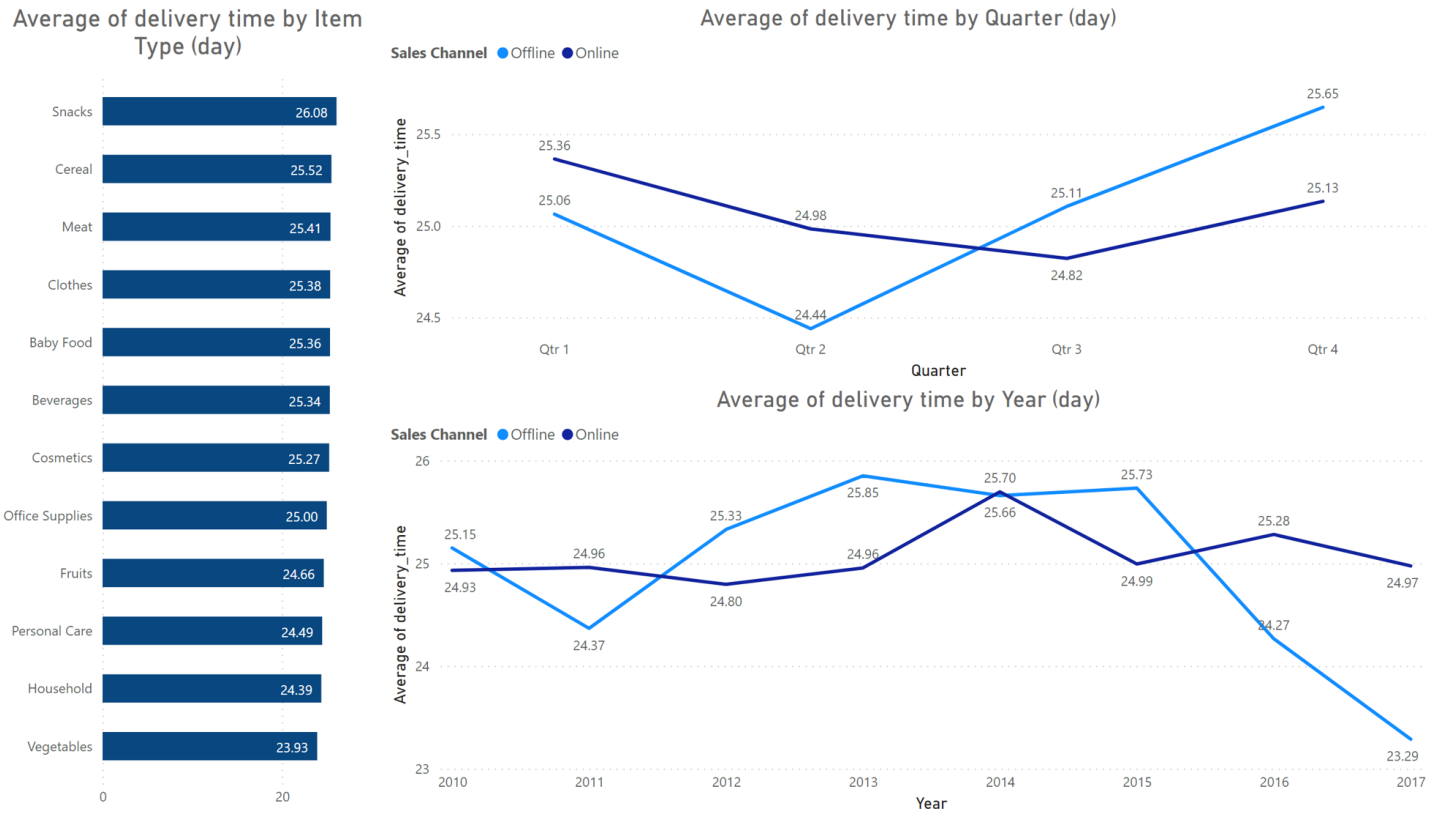
- Delivery time by country

Some countries have noticeable longer than average time, distributed mainly in Africa, Europe, Oceania.

(Map Average of delivery time by Country: the darker the color, the longer the delivery time)



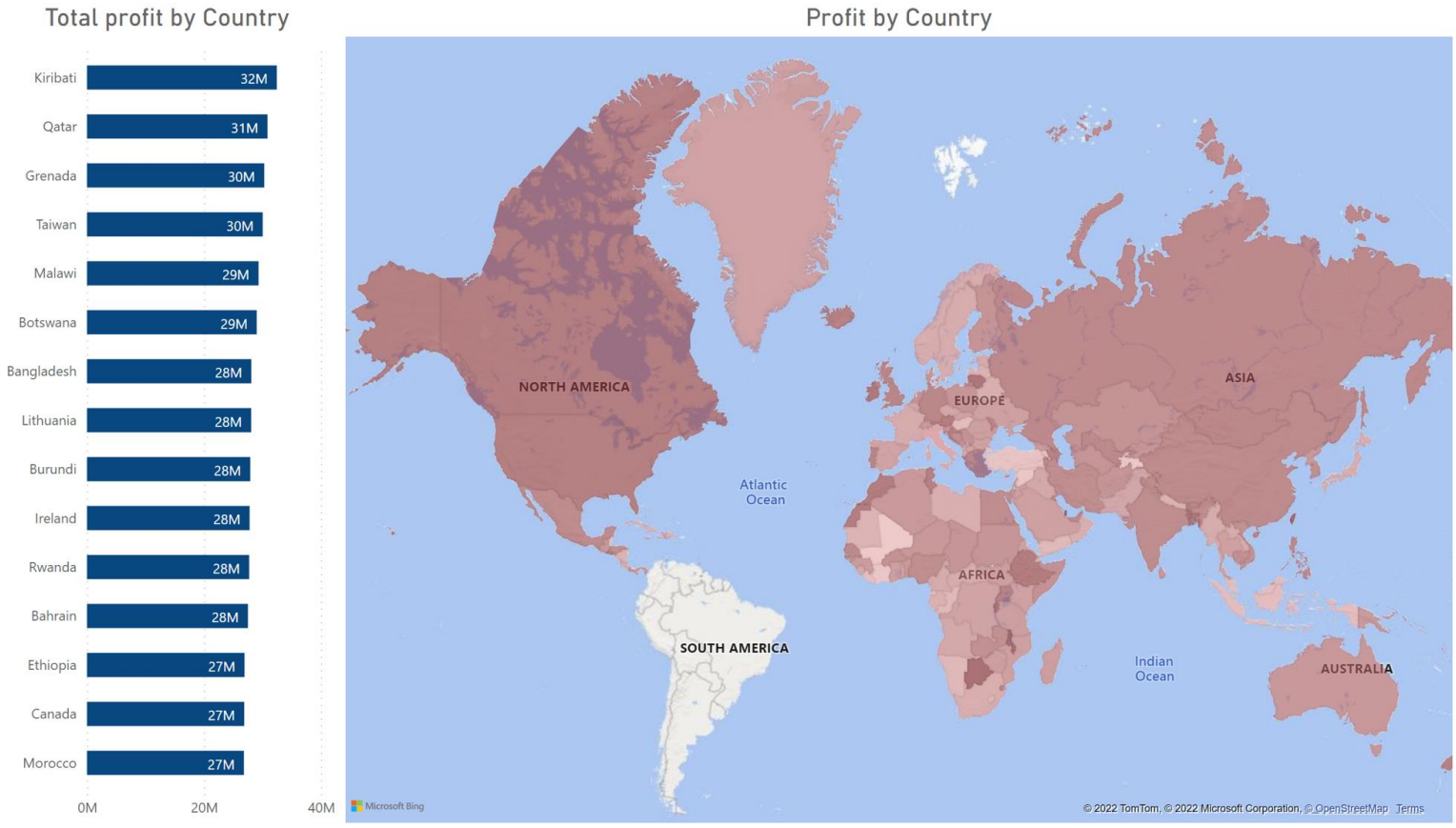
- Delivery time by item type and time



- Online channel has not changed much in delivery time through years, while offline has decreased since 2015. Offline chanel delivery time dropped noticably in Q2, increased in Q3, Q4, the difference between Q2 and Q4 is about 1 day. Online channel decreases slightly in Q2, Q3.
- For Item type, snacks, cereal, meat take longer to delivery than others. Meat should speed up delivery such as vegetables, fruits to ensure quality.

2. Profit, profit ratio, profit growth:

- Profit by Country
  (Map Profit by Country: the darker the color, the higer the total profit)

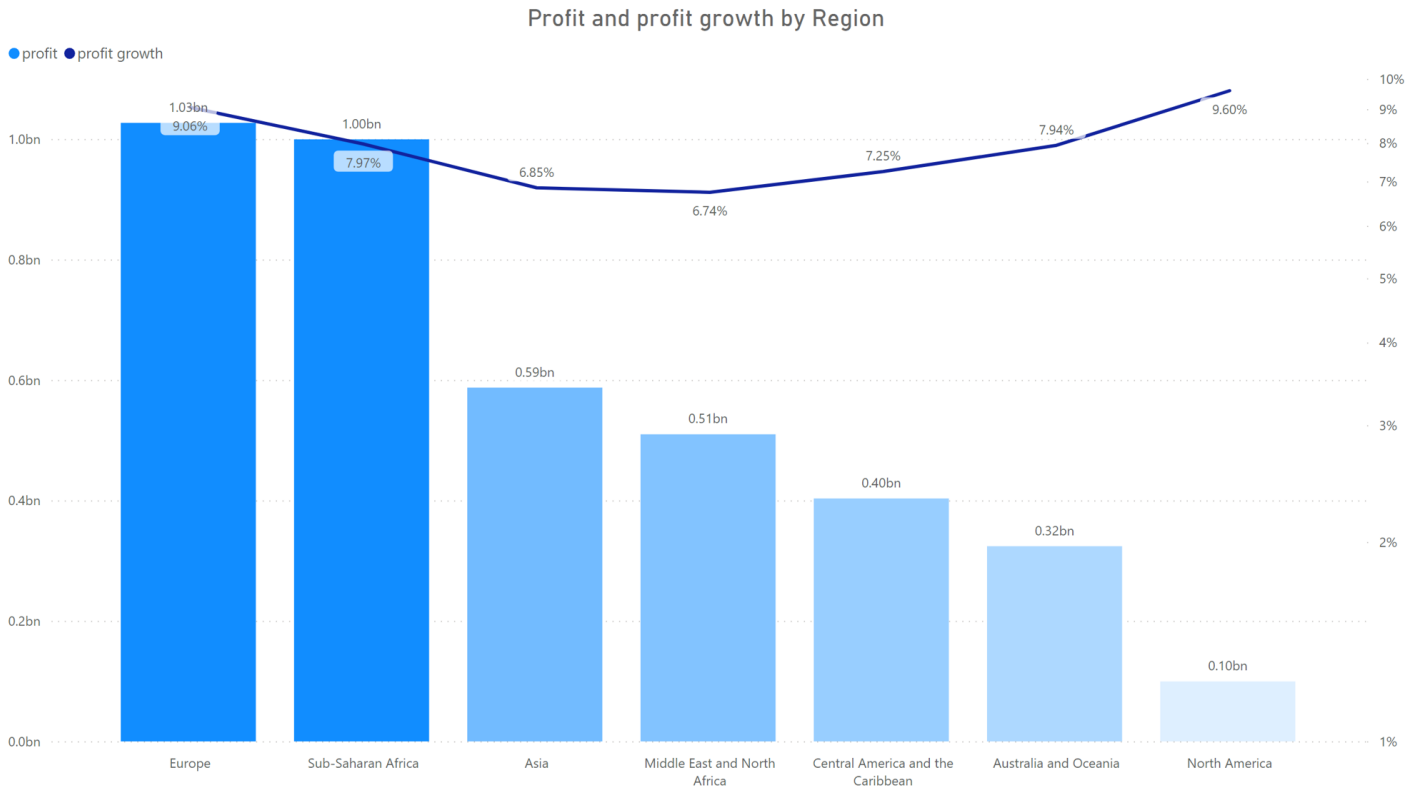


The profit source of the company is very large and diverse from all over the world except for South America. Some countries with small area in Africa, Europe bring relatively high profits for the company compared to big ones such as USA, Canada, China, Russia, Australia, India,…

- Profit, profit growth by region

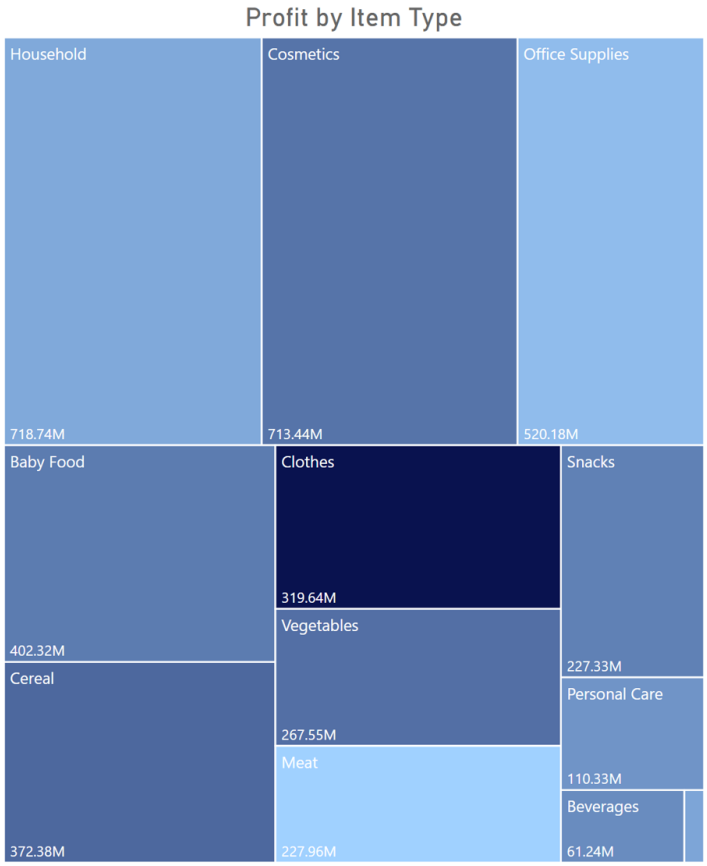
Europe and Sub-Saharan Africa sold a large number of products compared to other regions and are also two prominent regions contributing to the company's earnings. Europe also had a high profit growth index compared to others, but more notably North America, 9.6% growth over the years, perhaps the potential of this market needs to be exploited more strongly.

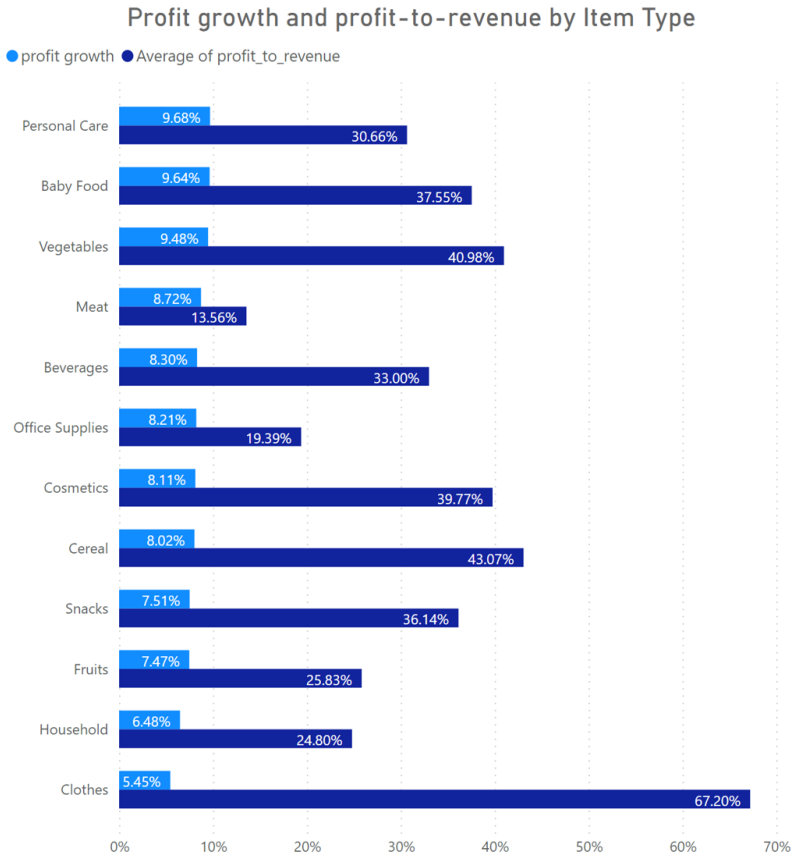
(The darker the color, the more total unit sold)



- Profit, profit growth, profit-to-revenue by item type

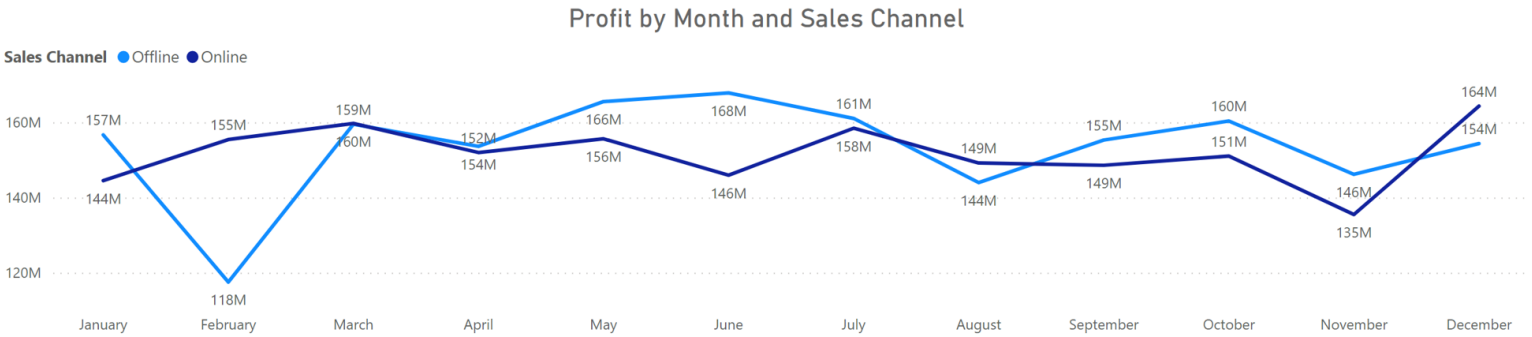
(Treemap Profit by Item Type: the darker the color, the higher profit_to_revenue)

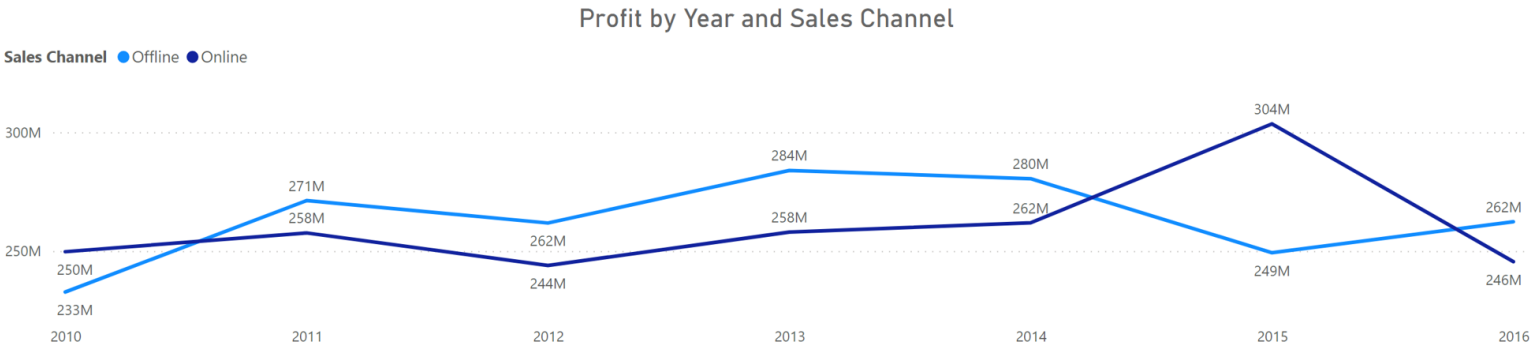


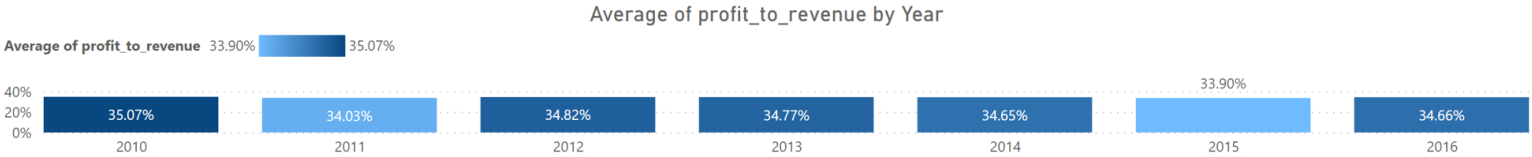


It can be seen that household, cosmetics, office supplies, baby food are the items that contribute great profits to the company. However, household and office supplies have a low profit-to-revenue ratio. Notably clothes, this is the most profitable item but the slowest growth rate over the years.

- Profit, profit-to-revenue by time, channel







- *Since the dataset does not have enough 12 months in 2017, it has been filtered out in these charts.*

Profit from offline channel is usually slightly higher. In general, they tend to increase from 2010 to 2014, but not strong growth over the years. In 2015, the offline channel decreased slightly, making up for the rise of the online channel, but dropped significantly in 2016. In 2015, 1 unit of revenue is not as profitable as before, but it was slightly restored in 2016.

In short, this chart shows that the company is in a slow development stage and is sometime backward. Therefore, it is necessary to study more deeply about other data such as customers, markets, products to have a strong growth strategy.