Questions to be answered

- How many sales have they made with amounts more than 1000
- How many sales have they made that belong to the Category "Tops" and have a Quantity of 3.
- The Total Sales by Category
- Average Amount by Category and Status
- Total Sales by Fulfilment and Shipment Type

Project 1 – Analyzing Amazon Sales Data

Importing the data

```
import pandas as pd

#Load the sales data from the excel file into a pandas DataFrame

sales_data = pd.read_excel('sales_data.xlsx')
```

Exploring the data

```
#get a summary of sales data
sales_data.info() #can also check data types here
sales_data.describe()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 128975 entries, 0 to 128974
Data columns (total 23 columns):
 #   Column              Non-Null Count   Dtype
---  ------              --------------   -----
 0   index               128975 non-null  int64
 1   Order ID            128975 non-null  object
 2   Date                128975 non-null  object
 3   Status              128975 non-null  object
 4   Fulfilment          128975 non-null  object
 5   Sales Channel       128975 non-null  object
 6   ship-service-level  128975 non-null  object
 7   Style               128975 non-null  object
 8   SKU                 128975 non-null  object
 9   Category            128975 non-null  object
 10  Size                128975 non-null  object
 11  ASIN                128975 non-null  object
 12  Courier Status      122103 non-null  object
 13  Qty                 128975 non-null  int64
 14  currency            121180 non-null  object
 15  Amount              121180 non-null  float64
 16  ship-city           128942 non-null  object
```

```
 17  ship-state          128935 non-null  object
 18  ship-postal-code    128935 non-null  float64
 19  ship-country        128935 non-null  object
 20  promotion-ids       79815 non-null   object
 21  B2B                 128968 non-null  float64
 22  fulfilled-by        39276 non-null   object
dtypes: float64(3), int64(2), object(18)
memory usage: 22.6+ MB
```

```
                index            Qty         Amount  ship-postal-
code  \
count  128975.000000  128975.000000  121180.000000     128935.000000

mean    64487.000000       0.904431     648.561465     463963.108256

std     37232.019822       0.313354     281.211687     191475.865632

min         0.000000       0.000000       0.000000     110001.000000

25%     32243.500000       1.000000     449.000000     382421.000000

50%     64487.000000       1.000000     605.000000     500033.000000

75%     96730.500000       1.000000     788.000000     600024.000000

max    128974.000000      15.000000    5584.000000     989898.000000


                 B2B
count  128968.000000
mean        0.006754
std         0.081903
min         0.000000
25%         0.000000
50%         0.000000
75%         0.000000
max         1.000000
```

```
#looking at columns
print(sales_data.columns)
```

```
Index(['index', 'Order ID', 'Date', 'Status', 'Fulfilment', 'Sales
Channel ',
       'ship-service-level', 'Style', 'SKU', 'Category', 'Size',
'ASIN',
       'Courier Status', 'Qty', 'currency', 'Amount', 'ship-city',
       'ship-state', 'ship-postal-code', 'ship-country', 'promotion-
ids',
       'B2B', 'fulfilled-by'],
      dtype='object')
```

```
#having a look at the first few rows of data
print(sales_data.head())
```

```
   index             Order ID      Date                          Status
\
0      0  405-8078784-5731545  04-30-22                       Cancelled

1      1  171-9198151-1101146  04-30-22  Shipped - Delivered to Buyer

2      2  404-0687676-7273146  04-30-22                         Shipped

3      3  403-9615377-8133951  04-30-22                       Cancelled

4      4  407-1069790-7240320  04-30-22                         Shipped


  Fulfilment Sales Channel  ship-service-level    Style
SKU  \
0   Merchant      Amazon.in           Standard   SET389   SET389-KR-
NP-S
1   Merchant      Amazon.in           Standard  JNE3781  JNE3781-KR-
XXXL
2     Amazon      Amazon.in          Expedited  JNE3371    JNE3371-KR-
XL
3   Merchant      Amazon.in           Standard    J0341       J0341-
DR-L
4     Amazon      Amazon.in          Expedited  JNE3671  JNE3671-TU-
XXXL

  Category  ... Qty currency  Amount    ship-city   ship-state  \
0      Set  ...   0      INR  647.62       MUMBAI  MAHARASHTRA
1   Blouse  ...   1      INR  406.00    BENGALURU    KARNATAKA
2   Blouse  ...   1      INR  329.00  NAVI MUMBAI  MAHARASHTRA
3    Dress  ...   0      INR  753.33   PUDUCHERRY   PUDUCHERRY
4      Top  ...   1      INR  574.00      CHENNAI   TAMIL NADU

   ship-postal-code ship-country  \
0          400081.0           IN
1          560085.0           IN
2          410210.0           IN
3          605008.0           IN
4          600073.0           IN

                                      promotion-ids  B2B fulfilled-by

0                                               NaN  0.0    Easy Ship

1  Amazon PLCC Free-Financing Universal Merchant ...  0.0    Easy Ship

2       IN Core Free Shipping 2015/04/08 23-48-5-108  1.0          NaN
```

```
3                                                   NaN  0.0    Easy Ship

4                                                   NaN  0.0          NaN


[5 rows x 23 columns]
```

```
#check the data types of the columns
print(sales_data.dtypes)
```

```
index                   int64
Order ID               object
Date                   object
Status                 object
Fulfilment             object
Sales Channel          object
ship-service-level     object
Style                  object
SKU                    object
Category               object
Size                   object
ASIN                   object
Courier Status         object
Qty                     int64
currency               object
Amount                float64
ship-city              object
ship-state             object
ship-postal-code      float64
ship-country           object
promotion-ids          object
B2B                   float64
fulfilled-by           object
dtype: object
```

Clearing the data

```
#Check for missing values in our sales data
print(sales_data.isnull().sum())
```

```
index                     0
Order ID                  0
Date                      0
Status                    0
Fulfilment                0
Sales Channel             0
ship-service-level        0
Style                     0
SKU                       0
Category                  0
```

```
Size                      0
ASIN                      0
Courier Status         6872
Qty                       0
currency               7795
Amount                 7795
ship-city                33
ship-state               40
ship-postal-code         40
ship-country             40
promotion-ids         49160
B2B                       7
fulfilled-by          89699
dtype: int64
```

```
#drop any rows that has any missing/nan values
sales_data_dropped = sales_data.dropna()

#drop rows with missing amounts based on the amount column
sales_data_cleaned = sales_data.dropna(subset = ['Amount'])

#Check for missing values in our sales data cleaned
print(sales_data_cleaned.isnull().sum())
```

```
index                     0
Order ID                  0
Date                      0
Status                    0
Fulfilment                0
Sales Channel             0
ship-service-level        0
Style                     0
SKU                       0
Category                  0
Size                      0
ASIN                      0
Courier Status         5136
Qty                       0
currency                  0
Amount                    0
ship-city                31
ship-state               38
ship-postal-code         38
ship-country             38
promotion-ids         41705
B2B                       7
fulfilled-by          83640
dtype: int64
```

Slicing and filtering the data

```
#Select a subset of our data based on the Category Column
category_data = sales_data[sales_data['Category'] == 'Top']
print(category_data)
```

```
         index             Order ID      Date
Status  \
4            4  407-1069790-7240320  04-30-22
Shipped
47          47  408-9281152-6213100  04-30-22
Shipped
54          54  402-7944191-1869101  04-30-22
Shipped
67          67  404-4535078-5241919  04-30-22
Shipped
109        109  403-7266170-1163500  04-30-22  Shipped - Delivered to
Buyer
...        ...                  ...       ...
...
128923  128923  402-1698543-5164342  05-31-22
Shipped
128941  128941  403-3035549-4142738  05-31-22
Shipped
128942  128942  403-3035549-4142738  05-31-22
Shipped
128949  128949  403-7050981-8813953  05-31-22
Shipped
128965  128965  408-5154281-4593912  05-31-22
Cancelled

       Fulfilment Sales Channel  ship-service-level    Style
SKU  \
4          Amazon      Amazon.in          Expedited  JNE3671  JNE3671-
TU-XXXL
47         Amazon      Amazon.in          Expedited  JNE3744
JNE3744-TU-S
54         Amazon      Amazon.in          Expedited    J0301
J0301-TP-L
67         Amazon      Amazon.in          Expedited  JNE3645
JNE3645-TP-N-L
109      Merchant      Amazon.in           Standard  JNE3648
JNE3648-TP-N-M
...           ...            ...                ...      ...
...
128923     Amazon      Amazon.in          Expedited    J0418
J0418-TP-XL
128941     Amazon      Amazon.in          Expedited    J0118
J0118-TP-XXL
128942     Amazon      Amazon.in          Expedited    J0301
```

```
J0301-TP-XXL
128949     Amazon      Amazon.in          Expedited    J0113
J0113-TP-M
128965     Amazon      Amazon.in          Expedited    J0119    J0119-
TP-XXXL

       Category  ... Qty currency Amount              ship-city  \
4           Top  ...   1      INR  574.0              CHENNAI
47          Top  ...   1      INR  665.0           SECUNDERABAD
54          Top  ...   1      INR  493.0              BENGALURU
67          Top  ...   1      INR  432.0                   PUNE
109         Top  ...   1      INR  518.0                 MUMBAI
...         ...  ...  ..      ...    ...                    ...
128923      Top  ...   1      INR  749.0              HYDERABAD
128941      Top  ...   1      INR  529.0              BENGALURU
128942      Top  ...   1      INR  518.0              BENGALURU
128949      Top  ...   1      INR  574.0              CHENNAI
128965      Top  ...   1      INR  574.0  Prayagraj (ALLAHABAD)

          ship-state  ship-postal-code ship-country  \
4          TAMIL NADU          600073.0           IN
47          TELANGANA          500017.0           IN
54          KARNATAKA          560076.0           IN
67        MAHARASHTRA          411014.0           IN
109       MAHARASHTRA          400076.0           IN
...               ...               ...          ...
128923      TELANGANA          500084.0           IN
128941      KARNATAKA          560102.0           IN
128942      KARNATAKA          560102.0           IN
128949     TAMIL NADU          600021.0           IN
128965  UTTAR PRADESH          211007.0           IN

                                           promotion-ids  B2B
fulfilled-by
4                                                    NaN  0.0
NaN
47           IN Core Free Shipping 2015/04/08 23-48-5-108  0.0
NaN
54           IN Core Free Shipping 2015/04/08 23-48-5-108  0.0
NaN
67           IN Core Free Shipping 2015/04/08 23-48-5-108  0.0
NaN
109     Amazon PLCC Free-Financing Universal Merchant ...  0.0    Easy
Ship
...                                                  ...  ...
...
128923  IN Core Free Shipping 2015/04/08 23-48-5-108,V...  0.0
NaN
128941                                               NaN  0.0
NaN
```

```
128942                                                    NaN  0.0
NaN
128949                                                    NaN  0.0
NaN
128965                                                    NaN  0.0
NaN

[10622 rows x 23 columns]
```

```
#Select a subset of our data where the Amount > 1000
high_amount_data = sales_data[sales_data['Amount'] > 1000]
print(high_amount_data)
```

```
         index             Order ID      Date
Status  \
32          32  404-9632124-1107550  04-30-22  Shipped - Delivered to
Buyer
43          43  408-3478480-0881162  04-30-22
Shipped
46          46  408-3917043-5314763  04-30-22
Shipped
52          52  408-7138000-9728362  04-30-22
Shipped
69          69  405-6480932-8759528  04-30-22
Shipped
...        ...                  ...       ...
...
128938  128938  403-6231612-3153120  05-31-22
Shipped
128954  128954  403-8464685-9792368  05-31-22
Shipped
128957  128957  402-8261465-0622733  05-31-22
Shipped
128966  128966  406-9812666-2474761  05-31-22
Shipped
128973  128973  402-6184140-0545956  05-31-22
Shipped

       Fulfilment Sales Channel  ship-service-level   Style
SKU  \
32       Merchant      Amazon.in           Standard   J0011
J0011-LCD-M
43         Amazon      Amazon.in          Expedited  SET360   SET360-
KR-NP-M
46         Amazon      Amazon.in          Expedited   J0230
J0230-SKD-L
52         Amazon      Amazon.in          Expedited  SET282   SET282-
KR-PP-M
69         Amazon      Amazon.in          Expedited   J0381   J0381-
SKD-XXXL
```

```
...           ...            ...                ...     ...
...
128938     Amazon      Amazon.in          Expedited   J0002
J0002-SKD-S
128954     Amazon      Amazon.in          Expedited   J0376
J0376-SKD-S
128957     Amazon      Amazon.in          Expedited   J0230
J0230-SKD-M
128966     Amazon      Amazon.in          Expedited  SET224  SET224-
KR-NP-XS
128973     Amazon      Amazon.in          Expedited   J0012     J0012-
SKD-XS

       Category  ... Qty currency  Amount       ship-city      ship-
state  \
32          Set  ...   1      INR  1233.0   VISAKHAPATNAM  ANDHRA
PRADESH
43          Set  ...   1      INR  1126.0         LUCKNOW   UTTAR
PRADESH
46          Set  ...   1      INR  1146.0          ONGOLE  ANDHRA
PRADESH
52          Set  ...   1      INR  1033.0         dimapur
NAGALAND
69          Set  ...   1      INR  1146.0        VIJAPURA
KARNATAKA
...         ...  ...  ..      ...     ...             ...
...
128938      Set  ...   1      INR  1186.0       GORAKHPUR   UTTAR
PRADESH
128954      Set  ...   1      INR  1063.0       NEW DELHI
DELHI
128957      Set  ...   1      INR  1349.0          RANCHI
JHARKHAND
128966      Set  ...   1      INR  1132.0  CHENNAI 600042      TAMIL
NADU
128973      Set  ...   1      INR  1199.0           Halol
Gujarat

        ship-postal-code ship-country  \
32              530016.0           IN
43              226017.0           IN
46              523001.0           IN
52              797112.0           IN
69              586201.0           IN
...                  ...          ...
128938          273006.0           IN
128954          110070.0           IN
128957          834002.0           IN
128966          600042.0           IN
```

```
128973          389350.0           IN

                                                promotion-ids  B2B
fulfilled-by
32      Amazon PLCC Free-Financing Universal Merchant ...  0.0    Easy
Ship
43           IN Core Free Shipping 2015/04/08 23-48-5-108  0.0
NaN
46                                                    NaN  0.0
NaN
52           IN Core Free Shipping 2015/04/08 23-48-5-108  0.0
NaN
69           IN Core Free Shipping 2015/04/08 23-48-5-108  0.0
NaN
...                                                   ...  ...
...
128938                                                NaN  0.0
NaN
128954                                                NaN  0.0
NaN
128957       IN Core Free Shipping 2015/04/08 23-48-5-108  0.0
NaN
128966                                                NaN  0.0
NaN
128973       IN Core Free Shipping 2015/04/08 23-48-5-108  0.0
NaN

[13332 rows x 23 columns]
```

```
#Select a subset of data based on multiple conditions
filtered_data = sales_data[(sales_data['Category'] == 'Top') &
(sales_data['Qty'] == 3)]
```

Aggregating the data

```
#total sales by category
category_totals = sales_data.groupby('Category')['Amount'].sum()
category_totals = sales_data.groupby('Category', as_index=False)
['Amount'].sum()
category_totals = category_totals.sort_values('Amount',
ascending=False)

#calculate the average Amount by Category and Fulfilment
fulfilment_averages = sales_data.groupby(['Category', 'Fulfilment'],
as_index=False)['Amount'].mean()
fulfilment_averages = fulfilment_averages.sort_values('Amount',
ascending=False)

#calculate the average Amount by Category and Status
status_averages = sales_data.groupby(['Category', 'Status'],
```

```
as_index=False)['Amount'].mean()
status_averages = status_averages.sort_values('Amount',
ascending=False)

#calculate total sales by shipment and fulfilment
total_sales_shipandfulfil = sales_data.groupby(['Courier Status',
'Fulfilment'], as_index=False)['Amount'].sum()
total_sales_shipandfulfil =
total_sales_shipandfulfil.sort_values('Amount', ascending=False)
```

Exporting the data

```
total_sales_shipandfulfil.rename(columns={'Courier Status' :
'Shipment'}, inplace = True)
status_averages.to_excel('average_sales_by_category_and_status.xlsx',
index=False)
total_sales_shipandfulfil.to_excel('total_sales_by_ship_and_fulfil.xls
x', index=False)
```