

E commerce (Lazada) Sale Analysis

Problem Statement

Lazada, one of the leading e-commerce platforms in Southeast Asia, generates a large amount of transactional and customer data every day. However, the company faces challenges in utilizing this data effectively to gain deep insights into customer purchasing behavior, identify sales trends, and measure the effectiveness of marketing strategies. Without proper analysis, Lazada may miss opportunities to improve customer engagement, optimize promotions, and increase overall revenue.

Objective:

To analyze customer purchasing behavior and sales trends on Lazada in order to identify key factors influencing customer decisions and overall platform performance.

Goals:

1. Identify the top-selling item categories and seasonal sales patterns.
2. Understand customer spending behavior.
3. Evaluate the effectiveness of Lazada's marketing campaigns and promotions.
4. Provide data-driven recommendations to improve customer engagement and increase revenue.
5. Shipping Fee effect to spending customers

Dataset

<https://www.lazada.co.th/#?>

EDA (Data Cleaning)

1. Dealing missing Value by drop

```
3
4 df_clean.dropna(axis=1,how='all',inplace=True)
5
```

```
3
4 df_clean.drop(columns='branch_number',inplace =True)
```

2 Fill null with reasonable values

```
1 df_clean['shipping_fee'].fillna(0,inplace=True)  
2 df_clean['variation'].str.contains('UNKNOWN').values
```

3. Dropping Duplicates

```
df_clean.duplicated()
```

4. Format Data Types

```
df_clean = df_clean.convert_dtypes()
```

Data transform to datetime

```
df_clean['create_dt'] = pd.to_datetime(df_clean['create_dt'],format='%Y-%m-%d %H:%M:%S')
```

Insight Business Question

- **Q1 Total Revenue**

```
1 #TODO: Total Revenue (paid_price) |  
2 total_revenue = df['paid_price'].sum()  
3  
4 print('Total Revenue between {} and {} = {:.2f} THB'.format(df['create_dt'].min(),df['create_dt'].max(),total_revenue))  
Total Revenue between 2020-10-01 01:18:37 and 2020-12-24 23:01:32 = 847,847.3399999999 THB
```

- **Q2 20 Top Spender**

```
3 revenue_by_cust_sorted = revenue_by_cust.sort_values(ascending = False)
4 revenue_by_cust_sorted.head(20)
```

	paid_price
	cust_name_hash
43b633560f95a64dbe4ab5eef390d280	15212.50
df64dc2eb4a0b85091dd31eb4923eaac	14825.30
eee14345482c566e7d72e988280be983	13460.75
372e7fd78fec6208c0ffab76b8e7b959	12199.00
7cd85ec5761a0856c803fff6396bb36a	12187.50
cc5abf089bcdd1b95c32c769aa296622	11668.75
980017891ff67cf8a20f23aa810e7b5a	11411.00
e8d8433622b25fb65e10a649b3c68d54	10892.00
f51d3f3feaf6d3d4e066ae03c78de189	10861.00
b5bd71e84bf5fdf530d4ce4597bb6ddc	10809.50
ec408f6e7f180cc0130d378c048c40db	9985.25
b441c1c7de2337e75694e53dab0727b0	9856.12
5f1557471819ecb2cb5b89526d30411f	8358.21
5075dc36e6fb5b2f19a38ac28d8a7fca	8274.22
7a34cf93564c34c6168dc2155b271944	8065.00
af381e406d4f95e2fa01cbd94f2f7b82	7929.22
7178a5ccd48ba326653d913b7d8ad433	7470.00
a1846b2ca04b92a2e357de285e6377fc	6940.00
fc69c355a6d6e138c78704ac12d0b14e	6935.00
af9f554e1aaaff11f3225dddb2660919	6440.00

- **Q3 Best Sell item category**

```
1 #TODO: Total Revenue in column paid_price (item_category)
2 df.groupby('item_category')['paid_price'].sum().to_frame('Total_sales')
```

Total_sales	
item_category	
ทินเนอร์	11748.08
น้ำมันสน	359.50
สีทาบ้าน	442262.29
สีทาฝ้าเพดาน	26488.82
สีน้ำมัน	3380.96
สีน้ำอะคริลิค	226302.19
สีรองพื้นกันสนิมเทา	105.00
สีรองพื้นปูนอเนกประสงค์	24140.38
สีรองพื้นปูนเก่า	93448.70
สีรองพื้นปูนใหม่กันด่าง	82686.04
อุปกรณ์อื่นๆ	1834.66

Scenario: Q4

Marketing team want to know insight data that what day of week has more transactions (0=Mon, 1=Tuesday, ..., 6=Sunday)

```
1 #Checking values of transactions
2 df['dow'].value_counts().sort_index()

→ count

dow
0    274
1    275
2    312
3    312
4    238
5    334
6    281
```

Q5 Checking order number are more than 1 shipping fee

```
1 #TODO: cheking order_no are more than 1 shipping fee
2 num_unique_fee_by_order[num_unique_fee_by_order>1]

→ shipping_fee

order_no
327845616538406      6
327873049142116      3
327878609492721      2
328005445690817      2
328050652120830      2
...
350717742799115      2
350798918951856      3
351415314281272      2
351452999463913      2
351470937948126      2
```

```

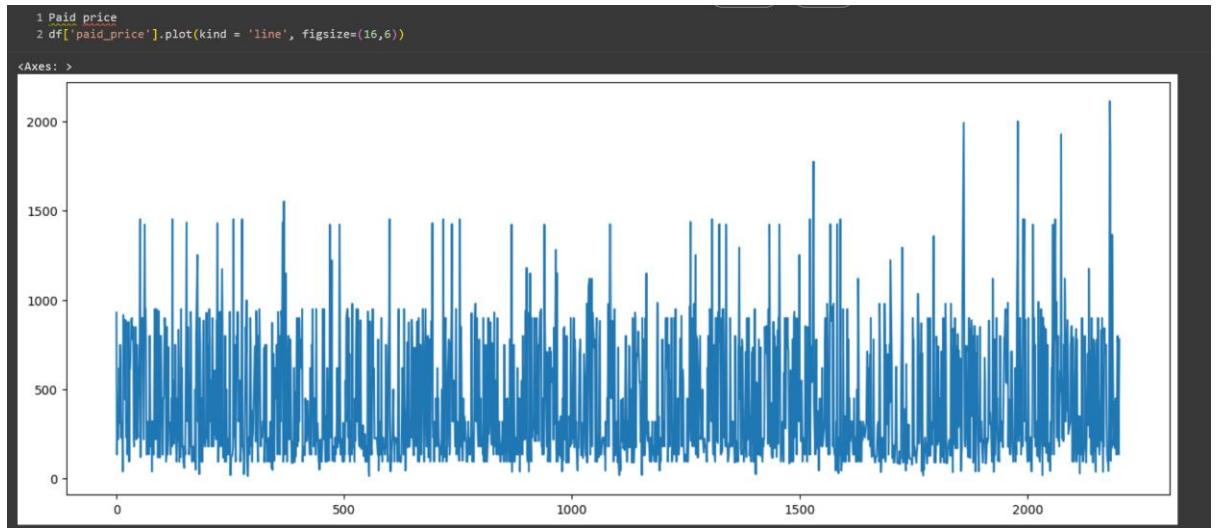
1 #TODO: select 1 order_no |
2
3 df[df['order_no'] == 327878609492721]

   order_no  lazada_sku  create_dt  cust_name_hash  shipping_name_hash  shipping_addr_hash  shipping_province  shipping_district  ...
25  327878609492721  1570134137_TH-4233738963  2020-10-01  a1846b2ca04b92a2e357de285e6377fc  1bc2f6e773ad72a6adbf4e9c46bb38f5  d958a0bbd09f5685161e206749cbbbae  อุบลฯ/ Sukhothai  วัดตัว/ Si Samrong
26  327878609492721  1024012823_TH-2267686441  2020-10-01  a1846b2ca04b92a2e357de285e6377fc  1bc2f6e773ad72a6adbf4e9c46bb38f5  d958a0bbd09f5685161e206749cbbbae  อุบลฯ/ Sukhothai  วัดตัว/ Si Samrong

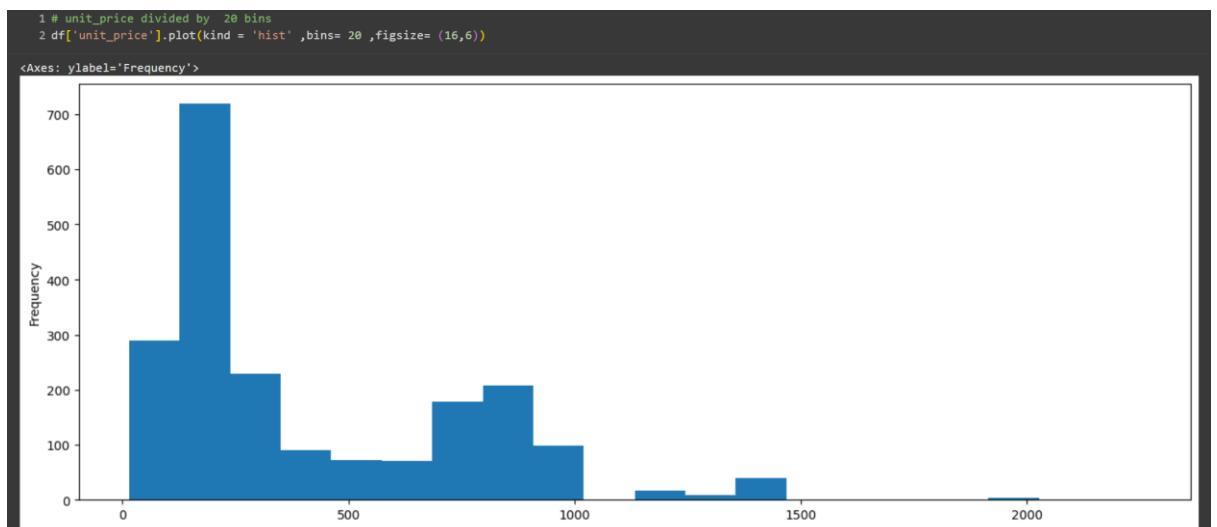
2 rows × 28 columns

```

Data Visualization

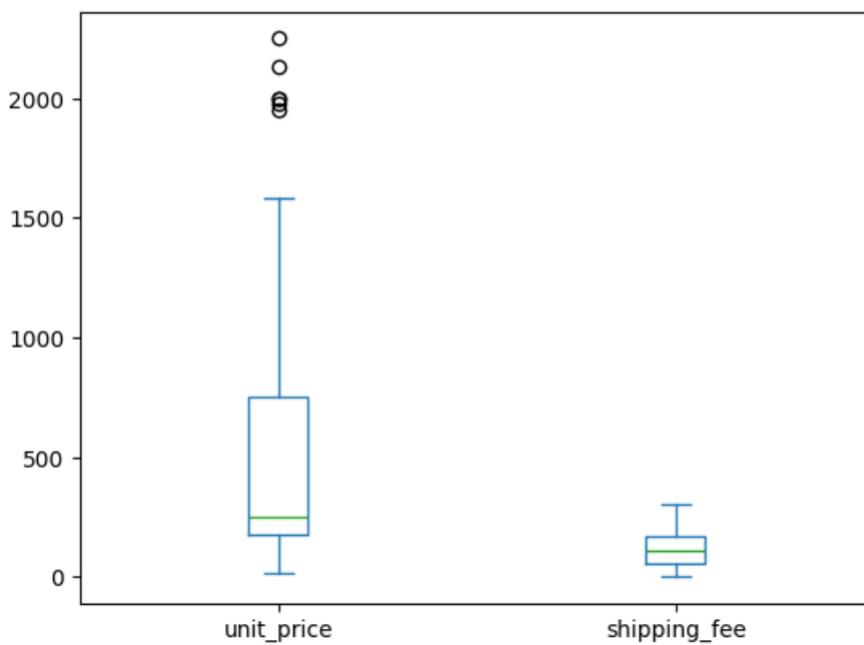


Unit_price



```
1 #TODO: Boxplot unit_price and shipping_fee
2 df[['unit_price','shipping_fee']].plot(kind = 'box')
```

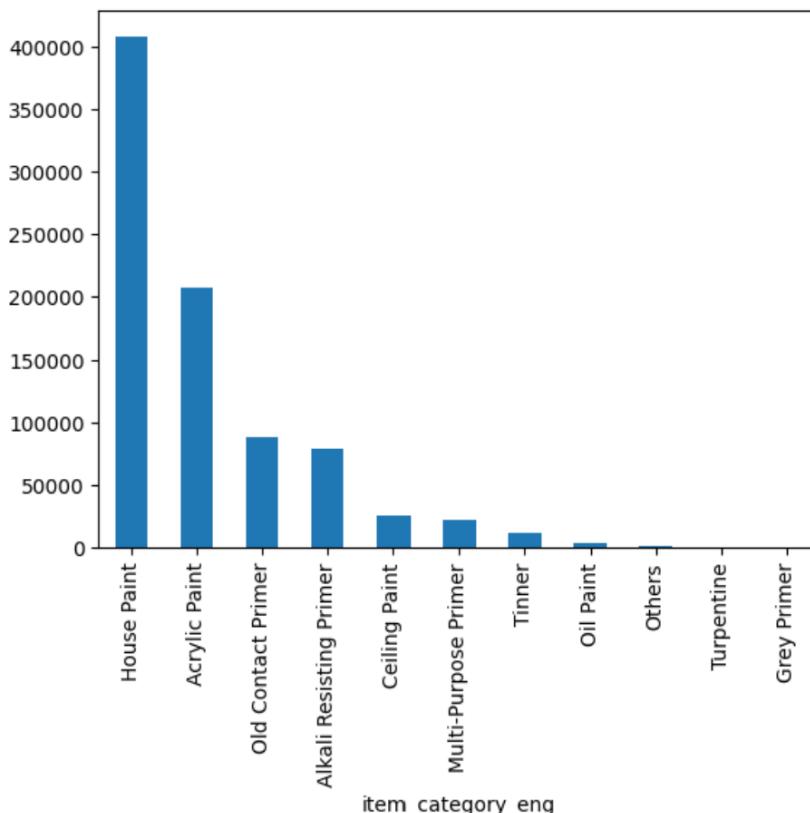
<Axes: >



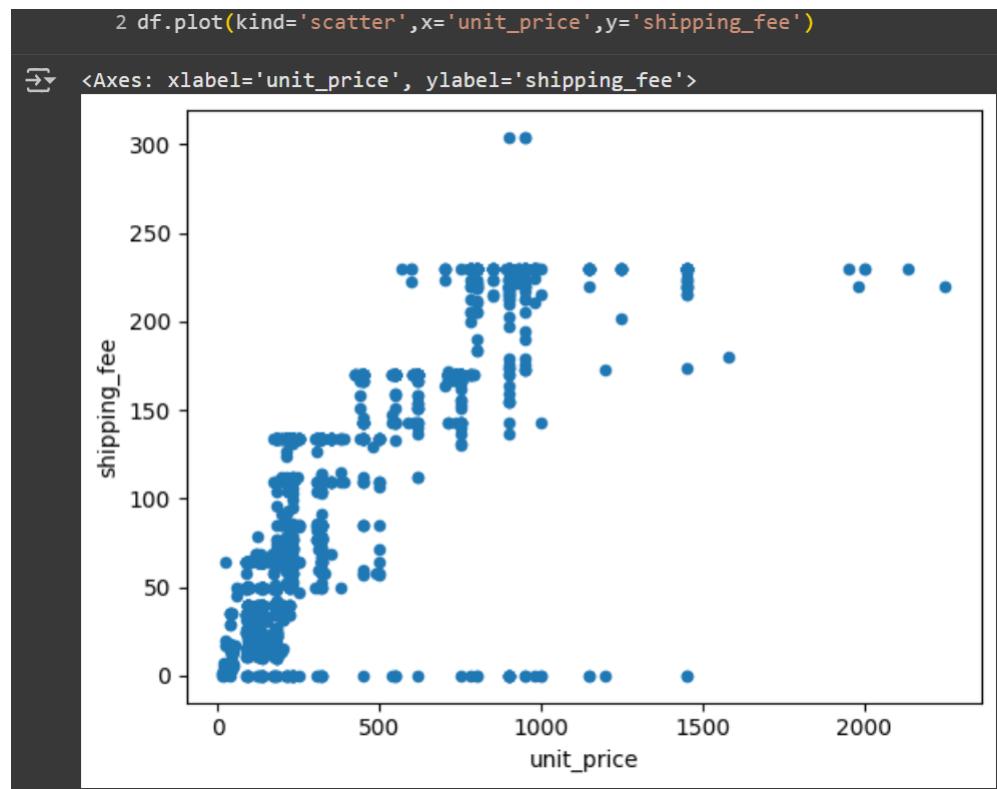
Item_category by Revenue

```
3 revenue_by_cat.sort_values(ascending=False).plot(kind='bar')
```

<Axes: xlabel='item_category_eng'>



Unit_price by shipping Fee



Daily Revenue

