Business context "HA" is a newly established firm operated in Viet Nam specializing in supplying various type of fruits for one of Vietnam largest distributor (will not be disclosed in this report due to privacy). For the scope of this project, timeline was between 01/10/2022 (the day the shop established) and 15/11/2022. Since this is a newly established shop, this project will only address two main question as following: How are drop-rates distributed across provinces, shops and products?

HA's Fruit Supply Business Report

Conducted by: Khang Le Nhu Gia

insufficient stocks or unqualified products).

Part 1: Revenue

Out[22]:

Which are the top performing provinces, shops and products regarding revenue? Answering these two question will help the firm have a accurate perspective about their main revenue generators as well as determine whether there is inefficiency in their operations (high drop-rate indicates

Q1: How was revenue distributed across provinces? sns.barplot(data = (df1.groupby('province')

.agg('sum') .filter(['province', 'revenue']) .sort values(by = 'revenue', ascending = False) .reset index()

.sort_values(by = 'revenue', ascending = False)

2.0

revenue

.sort values(by = 'revenue', ascending = False)

2.5

3.0

1e9

3.5 le9

color = 'darkcyan') <AxesSubplot:xlabel='province', ylabel='revenue'>

), x = 'province',y = 'revenue',

3.0 2.5

2.0

3.5

1.5

1.0 0.5

0.0

TP. Hổ Chí Minh TP. Đà Nẵng province

T. Bình Dương

TP. Đà Nẵng and TP. Hồ Chí Minh generated the lion share of total revenue (with revenue from TP. Đà Nẵng

almost tripled that from TP. Hồ Chí Minh). Later on, we will take a look at what are the main drivers (shops and products) of this trend.

Q2: How was revenue distributed across shops?

In [23]: sns.barplot(data = (df1.groupby('shop') .agg('sum') .filter(['shop', 'revenue'])

.reset_index()

x = 'revenue', color = 'darkcyan') <AxesSubplot:xlabel='revenue', ylabel='shop'>

> BDG Mỹ Phước 1 VCP HCM Ba Tháng Hai VCP HCM Lê Văn Việt HCM Bình Trưng HCM Hưng Gia VCC HCM Đồng Khởi VC+ HCM Phổ Quang

y = 'shop',

P_Fresh_Mién Trung_Trái cây Nhập P_Fresh_Miến Nam_Trái Cây Nhập Kh T_Fresh_Miến Nam_Trái Cây Nhập Kh T_Fresh_Mién Trung_Trái Cây Nhập VCC HCM Landmark 81

),

HCM Trung Sơn HCM Diamond VCC HCM Thảo Điển + HCM 26/4B Ấp Đông Lân VC+ HCM Nguyễn Duy Trinh

0.0

.agg('sum')

.reset index()),

1.0

The highest amount of revenue was generated by "P_Fresh_Miền Trung", a store from TP. Đà Nẵng; ranked in second is "P_Fresh Miền Nam" from TP. Hồ Chí Minh. "T_Fresh_Miền Trung" which ranked at fourth is from

TP. Đà Nẵng while the remaining stores are from either TP. Hồ Chí Minh or Bình Dương. Q3: How was revenue distributed across products?

sns.barplot(data = (df1.groupby('product')

<AxesSubplot:xlabel='revenue', ylabel='product'>

y = 'product', x = 'revenue', color = 'darkcyan')

Nho đen Mỹ

Táo Fuji Nam Phi Táo Koru New Zealand

Nho đỏ không hạt Mỹ Nho xanh Hàn Quốc Muscat

Lê Hàn Quốc

0.0

In [24]:

Out[24]:

Lê đỏ Nam phi Nho xanh không hat Mỹ Táo Gala Mỹ Táo Fuji Nam Phi túi 3kg

revenue

shop

P_Fresh_Miến Trung_Trái cây Nhập

P_Fresh_Miến Nam_Trái Cây Nhập Kh

There were two products that generated the most of total revenue: the first was "Nho đen Mỹ"; the second

Táo Fuji Nam Phi túi 3kg Nho đỏ không hạt Mỹ 0.2 0.4 1.0 1.2 0.0 0.6 0.8 1e9 revenue Interestingly, "Nho đen Mỹ" was supplied to only "P_Fresh_Miền Trung" from TP. Đà Nẵng not "P_Fresh_Miền Nam" from TP. Hồ Chí Minh. Despite this, "Nho đen Mỹ" contributed the most to total revenue. Part 2: Drop-rate Q5: How was drop-rate distributed? sns.histplot(data = df1, x = 'drop rate', bins = 30, color = 'darkred')In [27]: <AxesSubplot:xlabel='drop rate', ylabel='Count'> Out[27]:

0.8

.map(sns.histplot, 'drop rate', bins = 30, color = 'darkred'))

Drop-rate varied largely from 0%-20%, which was around the regular benchmark of the fruit industry.

However, there was a spike in the number of orders that had 90%-100% drop-rate; we will discover why this

province = TP. Đà Nẵng

0.0

0.5

roduct = Táo Fuji Nam Phi túi 3kg

It is now clear that "Nho đen Mỹ", "Nho đỏ không hạt Mỹ", "Nho xanh không hạt Mỹ" were the products

According to the operations team, for "Nho đen Mỹ", "Nho đỏ không hạt Mỹ", there was a problem with the

freezer leading to a high number of products being spoiled; for "Nho đỏ không hạt Mỹ" despite high

Despite having high drop-rate, "Nho đen Mỹ" was ranked first and "Nho xanh không hạt Mỹ" was ranked

drop_rate

1.0

province = T. Bình Dương

0.5

1.0

0.6 0.4

Ngà

hàr

442.32

price amount_ordered amount_delivered

6125.0

đ

1.0

0.6

-0.5 0.5 1.0 -0.5drop_rate Drilling the data to provinces we can see that the majority of orders that had drop-rates between 90%-100% came from TP. Đà Nẵng. We will continue drilling down to product-level in TP. Đà Nẵng to determine which

(sns.FacetGrid(data = df1.query('province == "TP. Đà Nẵng"'), col = 'product', col wrap In [29]: .map(sns.histplot, 'drop rate', bins = 30, color = 'darkred')) <seaborn.axisgrid.FacetGrid at 0x197ba78>

products caused this.

Out[30]: 5000 4000 amount delivered 3000 2000 1000 0

> 1000 2000 3000 4000 5000 amount_ordered In TP. Hồ Chí Minh, most orders were delivered; there was no horizontal line of dots; however the dot at around (6000,450) we seen in the previous report appears. Drill down to have a closer look at the peculiar order: dfl.query('province == "TP. Hồ Chí Minh" and amount ordered > 6000')

<AxesSubplot:xlabel='price', ylabel='drop_rate'>

sns.regplot(data = df1, x = 'price', y = 'drop_rate', color = 'darkred')

drop_rate 9.0 9.0 0.4 0.2 0.0 100000 200000 300000 400000 500000 price

drop_rate 0.2 0.0 -0.2

The distribution of revenue skewed heavily to shops in TP. Đà Nẵng which sold products such as "Nho đen Mỹ" and "Nho xanh không hạt Mỹ" although the drop-rate of these product were quite high. A fast actionable advice for the firm is that they should focus more on getting more volumn of these products and ensure that its quality meet the distributor's standard.

was "Lê Đỏ Nam Phi". Q4: Regarding the two highest revenue generating shops, which products were their main revenue drivers? sns.barplot(data = (dfl.query('shop == "P Fresh Miền Trung Trái cây Nhập" or shop == "P In [26]: .groupby(['shop','product']) .agg('sum') .sort_values(by = 'revenue', ascending = False) .reset index()), y = 'product', x = 'revenue', hue = 'shop',) <AxesSubplot:xlabel='revenue', ylabel='product'> Out[26]:

Nho đen Mỹ

Táo Gala Mỹ

Lê Hàn Quốc

Táo Fuji Nam Phi

Tão Koru New Zealand

Lê đỏ Nam phi

Nho xanh không hạt Mỹ

product

60

50

40

20

happen later on.

50

10

0.0

17.5 15.0 12.5 0.0 ti 5.0 2.5

product = Lê Hàn Quốc

third in terms of total revenue.

sns.regplot(data = df1,

1000

delivered; however there are two anomalies:

we have addressed earlier on)

A dot at roughly (6000,450)

In [31]:

Out[31]:

Out[32]:

In [33]:

Out[33]:

In [34]:

Out[34]:

0.8

0.6

0.4

1.4

1.2

1.0

2000

Count 30

In [28]:

Out[28]:

Out[29]:

10 0 0.0 0.2 0.4 drop_rate

40 30 20

Q6: How was drop-rate distributed across provinces?

<seaborn.axisgrid.FacetGrid at 0x599fb98>

province = TP. Hó Chí Minh

(sns.FacetGrid(data = df1, col = 'province')

15.0 12.5 10.0 7.5 5.0

that largely responsible for the 90-100% drop rate.

number of orders, this product was out-of-stock for the most part.

x = 'amount ordered',

color = 'darkred')

'amount delivered',

Q8: How was amount_ordered and amount_delivered correlate to one another?

<AxesSubplot:xlabel='amount ordered', ylabel='amount delivered'>

4000

3000

amount ordered

5000

amount_ordered and amount_delivered. The upward sloping line indicates that most orders were; in fact,

Horizontal line of dots at amount_delivered = 0 indicates orders that have 90-100% drop-rate (which

In order to see if there were more anomalies in drop-rate, we will look at the relationship between

Q8: In TP. Hồ Chí Minh, how were amount_ordered and amount_delivered correlate to one another?

sns.regplot(data = df1.query('province == "TP. Hô Chí Minh"'),

x = 'amount ordered',y = 'amount delivered', 6000

product = Táo Fuji Nam Phi

Q7: How was drop-rate distributed across products sold in TP. Đà Nẵng?

color = 'darkred') <AxesSubplot:xlabel='amount ordered', ylabel='amount delivered'> 6000 5000 amount delivered 4000 3000 2000 1000 0 shop product measurement po province P_Fresh_Miền Táo Fuji TP. Hồ **105** 4143856841 KG 50000 Nam_Trái Cây Nam Chí Minh Phi Nhập Kh This order, according to management, was not delivered as ordered because of insufficient stock. Q9: Besides product's quality and availability, did price affect drop-rate?

The relationship between price and drop-rate was quite abitrary suggesting drop-rate is more productbased than price-based Q10: Besides product's quality and availability, did ordered quantity affect drop-rate? sns.regplot(data = df1, x = 'amount_ordered', y = 'drop_rate', color = 'darkred') <AxesSubplot:xlabel='amount ordered', ylabel='drop rate'>

There was very little to no correlation between the amount ordered and drop-rate

5000 6000 2000 4000 3000 amount_ordered Part 3: Conclusion

1000