lcxyjjxbc

February 28, 2025

[3]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import datetime as dt
import warnings
warnings.filterwarnings('ignore')
import os
```

[9]:
```
df=pd.read_csv(r'C:\Users\kmuba\Downloads\index.csv')
```

[11]:
```
df
```

[11]:

	date	datetime	cash_type	card	\
0	2024-03-01	2024-03-01 10:15:50.520	card	ANON-0000-0000-0001	
1	2024-03-01	2024-03-01 12:19:22.539	card	ANON-0000-0000-0002	
2	2024-03-01	2024-03-01 12:20:18.089	card	ANON-0000-0000-0002	
3	2024-03-01	2024-03-01 13:46:33.006	card	ANON-0000-0000-0003	
4	2024-03-01	2024-03-01 13:48:14.626	card	ANON-0000-0000-0004	
…	…	…	…	…	
1128	2024-07-31	2024-07-31 20:53:35.077	card	ANON-0000-0000-0443	
1129	2024-07-31	2024-07-31 20:59:25.013	card	ANON-0000-0000-0040	
1130	2024-07-31	2024-07-31 21:26:26.000	card	ANON-0000-0000-0444	
1131	2024-07-31	2024-07-31 21:54:11.824	card	ANON-0000-0000-0445	
1132	2024-07-31	2024-07-31 21:55:16.570	card	ANON-0000-0000-0446	

	money	coffee_name
0	38.70	Latte
1	38.70	Hot Chocolate
2	38.70	Hot Chocolate
3	28.90	Americano
4	38.70	Latte
…	…	…
1128	23.02	Cortado
1129	27.92	Americano with Milk
1130	32.82	Latte
1131	32.82	Latte

```
1132  32.82                Latte

[1133 rows x 6 columns]
```

[13]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1133 entries, 0 to 1132
Data columns (total 6 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   date         1133 non-null   object
 1   datetime     1133 non-null   object
 2   cash_type    1133 non-null   object
 3   card         1044 non-null   object
 4   money        1133 non-null   float64
 5   coffee_name  1133 non-null   object
dtypes: float64(1), object(5)
memory usage: 53.2+ KB
```

[15]:
```
df.describe()
```

[15]:
```
              money
count  1133.000000
mean     33.105808
std       5.035366
min      18.120000
25%      28.900000
50%      32.820000
75%      37.720000
max      40.000000
```

[17]:
```
df.isnull().sum()
```

[17]:
```
date            0
datetime        0
cash_type       0
card           89
money           0
coffee_name     0
dtype: int64
```

[19]:
```
df.duplicated().sum()
```

[19]:
```
0
```

[21]:
```
df.describe().T
```

[21]:

	count	mean	std	min	25%	50%	75%	max
money	1133.0	33.105808	5.035366	18.12	28.9	32.82	37.72	40.0

[23]:
```
df.loc[:,['cash_type','card','coffee_name']].describe(
).T
```

[23]:

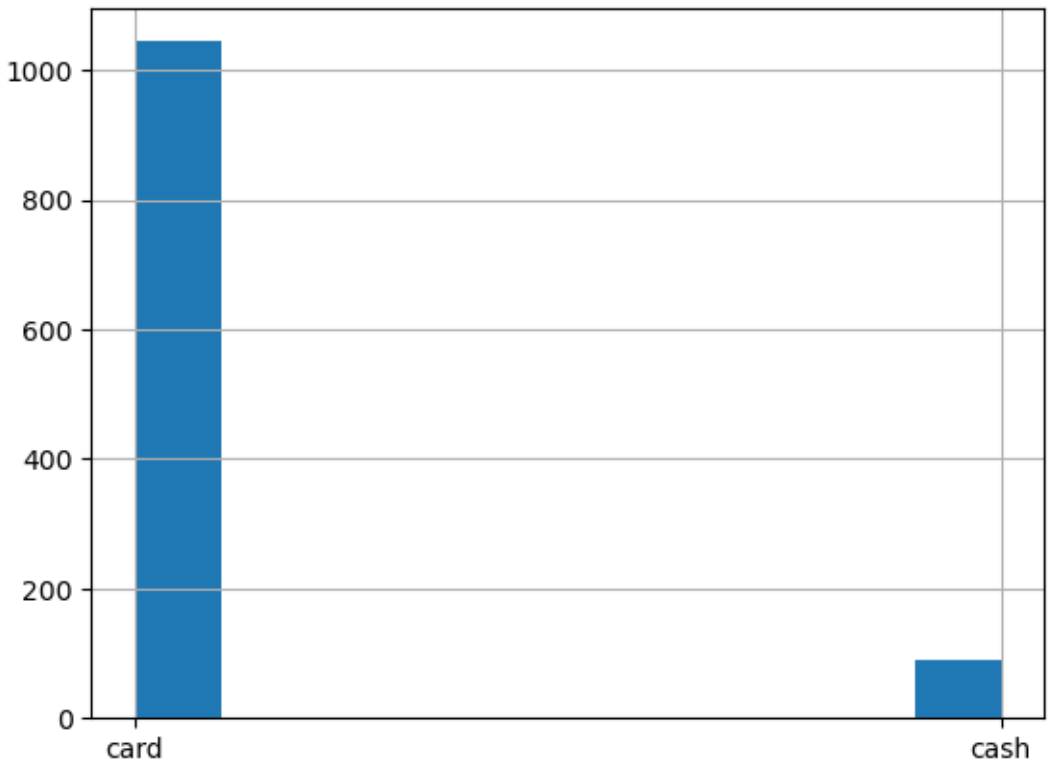
	count	unique	top	freq
cash_type	1133	2	card	1044
card	1044	446	ANON-0000-0000-0012	88
coffee_name	1133	8	Americano with Milk	268

[29]:
```
df[df['card'].isnull()]['cash_type'].value_counts()
```

[29]:
```
cash_type
cash    89
Name: count, dtype: int64
```

[31]:
```
df['cash_type'].hist()
```

[31]: <Axes: >



[33]:
```
df['cash_type'].value_counts(normalize=True)
```

[33]:
```
cash_type
card    0.921447
cash    0.078553
Name: proportion, dtype: float64
```

[35]:
```
pd.DataFrame(df['coffee_name'].value_counts(normalize=
True).sort_values(ascending=False).round(4)*100)
```

[35]:

coffee_name	proportion
Americano with Milk	23.65
Latte	21.45
Cappuccino	17.30
Americano	14.92
Cortado	8.74
Hot Chocolate	6.53
Espresso	4.32
Cocoa	3.09

[39]:
```
df['date']=pd.to_datetime(df['date'])
df['datetime']=pd.to_datetime(df['datetime'])
df['month']=df['date'].dt.strftime('%Y-%m')
df['day']=df['date'].dt.strftime('%w')
df['hour']=df['datetime'].dt.strftime('%H')
```

[41]:
```
df
```

[41]:

	date	datetime	cash_type	card	money	\
0	2024-03-01	2024-03-01 10:15:50.520	card	ANON-0000-0000-0001	38.70	
1	2024-03-01	2024-03-01 12:19:22.539	card	ANON-0000-0000-0002	38.70	
2	2024-03-01	2024-03-01 12:20:18.089	card	ANON-0000-0000-0002	38.70	
3	2024-03-01	2024-03-01 13:46:33.006	card	ANON-0000-0000-0003	28.90	
4	2024-03-01	2024-03-01 13:48:14.626	card	ANON-0000-0000-0004	38.70	
…	…	…	…	…	…	
1128	2024-07-31	2024-07-31 20:53:35.077	card	ANON-0000-0000-0443	23.02	
1129	2024-07-31	2024-07-31 20:59:25.013	card	ANON-0000-0000-0040	27.92	
1130	2024-07-31	2024-07-31 21:26:26.000	card	ANON-0000-0000-0444	32.82	
1131	2024-07-31	2024-07-31 21:54:11.824	card	ANON-0000-0000-0445	32.82	
1132	2024-07-31	2024-07-31 21:55:16.570	card	ANON-0000-0000-0446	32.82	

	coffee_name	month	day	hour
0	Latte	2024-03	5	10
1	Hot Chocolate	2024-03	5	12
2	Hot Chocolate	2024-03	5	12
3	Americano	2024-03	5	13
4	Latte	2024-03	5	13
…	…	…	..	…

```
1128              Cortado  2024-07   3   20
1129  Americano with Milk  2024-07   3   20
1130                Latte  2024-07   3   21
1131                Latte  2024-07   3   21
1132                Latte  2024-07   3   21

[1133 rows x 9 columns]
```

[43]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1133 entries, 0 to 1132
Data columns (total 9 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   date         1133 non-null   datetime64[ns]
 1   datetime     1133 non-null   datetime64[ns]
 2   cash_type    1133 non-null   object
 3   card         1044 non-null   object
 4   money        1133 non-null   float64
 5   coffee_name  1133 non-null   object
 6   month        1133 non-null   object
 7   day          1133 non-null   object
 8   hour         1133 non-null   object
dtypes: datetime64[ns](2), float64(1), object(6)
memory usage: 79.8+ KB
```

[45]:
```
[df['date'].min(),df['date'].max()]
```

[45]:
```
[Timestamp('2024-03-01 00:00:00'), Timestamp('2024-07-31 00:00:00')]
```

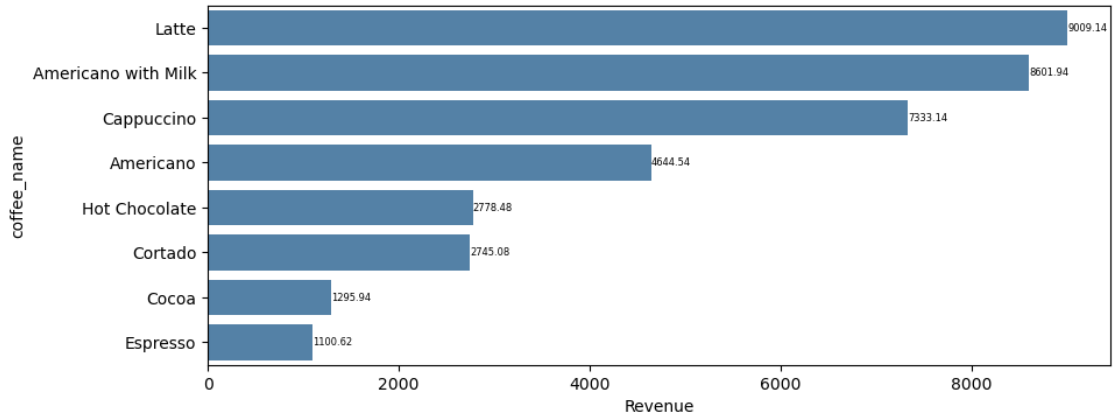
[53]:
```
revenue_data=df.groupby(['coffee_name']).sum(['money']).reset_index().
↪sort_values(by='money',ascending=False)
```

[55]:
```
revenue_data
```

[55]:
```
           coffee_name    money
7                Latte  9009.14
1  Americano with Milk  8601.94
2            Cappuccino  7333.14
0             Americano  4644.54
6         Hot Chocolate  2778.48
4               Cortado  2745.08
3                 Cocoa  1295.94
5              Espresso  1100.62
```

[57]:
```
plt.figure(figsize=(10,4))
ax =sns.barplot(data=revenue_data,x='money',y='coffee_name',color='steelblue')
ax.bar_label(ax.containers[0], fontsize=6)
plt.xlabel('Revenue')
```

[57]:
```
Text(0.5, 0, 'Revenue')
```



[63]:
```
monthly_sales =df.groupby(['coffee_name','month']).count()['date'].
↪reset_index().rename(columns={'date':'count'}).
↪pivot(index='month',columns='coffee_name',values='count').reset_index()
```

[65]:
```
monthly_sales
```

[65]:
```
coffee_name    month  Americano  Americano with Milk  Cappuccino  Cocoa  \
0            2024-03         36                   34          20      6
1            2024-04         35                   42          43      6
2            2024-05         48                   58          55      9
3            2024-06         14                   69          46      5
4            2024-07         36                   65          32      9

coffee_name  Cortado  Espresso  Hot Chocolate  Latte
0                 30        10             22     48
1                 19         7             13     31
2                 17         8             14     58
3                 19        10             14     50
4                 14        14             11     56
```

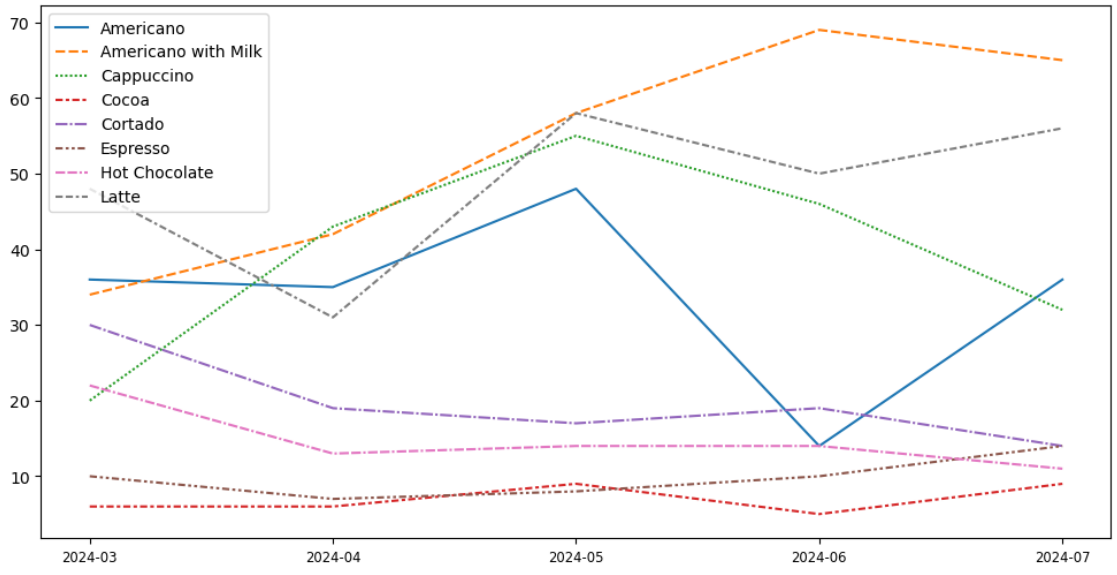
[67]:
```
monthly_sales.describe().T.loc[:,['min','max']]
```

[67]:
```
                     min   max
coffee_name
Americano           14.0  48.0
```

```
Americano with Milk  34.0  69.0
Cappuccino           20.0  55.0
Cocoa                 5.0   9.0
Cortado              14.0  30.0
Espresso              7.0  14.0
Hot Chocolate        11.0  22.0
Latte                31.0  58.0
```

[69]:
```
plt.figure(figsize=(12,6))
sns.lineplot(data=monthly_sales)
plt.legend(loc='upper left')
plt.
  ↪xticks(range(len(monthly_sales['month'])),monthly_sales['month'],size='small')
```

[69]:
```
([<matplotlib.axis.XTick at 0x24ca1e3b6e0>,
  <matplotlib.axis.XTick at 0x24ca1e3a990>,
  <matplotlib.axis.XTick at 0x24ca1e39cd0>,
  <matplotlib.axis.XTick at 0x24ca28ef0b0>,
  <matplotlib.axis.XTick at 0x24ca2962450>],
 [Text(0, 0, '2024-03'),
  Text(1, 0, '2024-04'),
  Text(2, 0, '2024-05'),
  Text(3, 0, '2024-06'),
  Text(4, 0, '2024-07')])
```



[75]:
```
weekday_sales =df.groupby(['day']).count()['date'].reset_index().
  ↪rename(columns={'date':'count'})
```

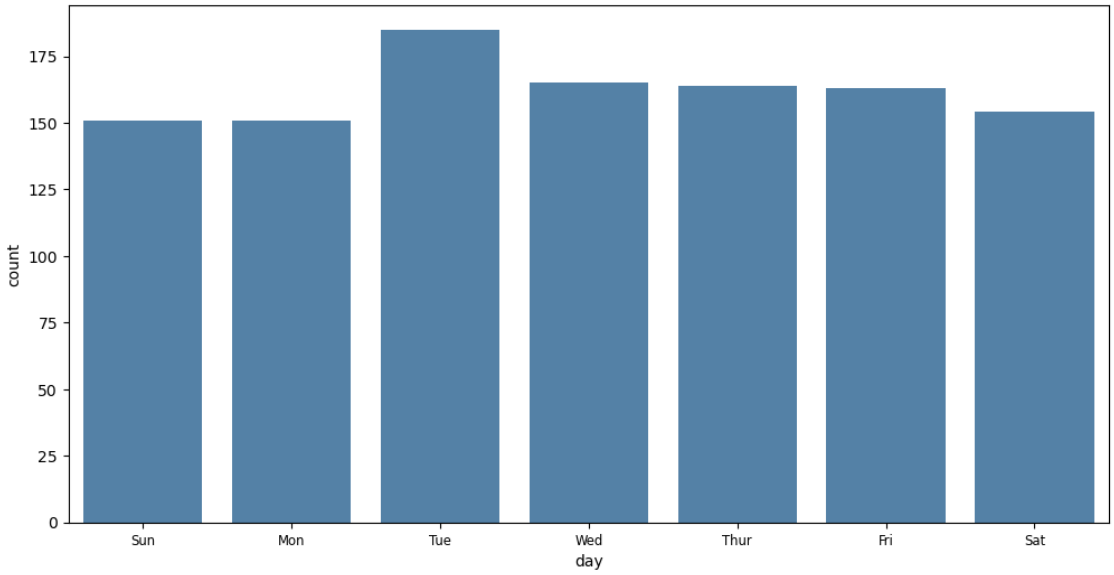
[77]:
```
weekday_sales
```

[77]:

	day	count
0	0	151
1	1	151
2	2	185
3	3	165
4	4	164
5	5	163
6	6	154

[79]:
```
plt.figure(figsize=(12,6))
sns.barplot(data=weekday_sales,x='day',y='count',color='steelblue')
plt.
  ↪xticks(range(len(weekday_sales['day'])),['Sun','Mon','Tue','Wed','Thur','Fri','Sat'],size='
```

[79]:
```
([<matplotlib.axis.XTick at 0x24ca27d8b60>,
  <matplotlib.axis.XTick at 0x24ca27d8830>,
  <matplotlib.axis.XTick at 0x24ca1ecf620>,
  <matplotlib.axis.XTick at 0x24ca203f0b0>,
  <matplotlib.axis.XTick at 0x24ca28187a0>,
  <matplotlib.axis.XTick at 0x24ca2819130>,
  <matplotlib.axis.XTick at 0x24ca2819a30>],
 [Text(0, 0, 'Sun'),
  Text(1, 0, 'Mon'),
  Text(2, 0, 'Tue'),
  Text(3, 0, 'Wed'),
  Text(4, 0, 'Thur'),
  Text(5, 0, 'Fri'),
  Text(6, 0, 'Sat')])
```



[83]:
```
daily_sales =df.groupby(['coffee_name','date']).count()['datetime'].
↪reset_index().reset_index().rename(columns={'datetime':'count'}).
↪pivot(index='date',columns='coffee_name',values='count').reset_index().
↪fillna(0)
```

[85]:
```
daily_sales
```

[85]:
```
coffee_name       date  Americano  Americano with Milk  Cappuccino  Cocoa  \
0            2024-03-01        1.0                  4.0         0.0    1.0
1            2024-03-02        3.0                  3.0         0.0    0.0
2            2024-03-03        1.0                  2.0         0.0    1.0
3            2024-03-04        0.0                  1.0         0.0    0.0
4            2024-03-05        0.0                  0.0         0.0    1.0
..                  …        …                    …          …     …
145          2024-07-27        0.0                  5.0         4.0    0.0
146          2024-07-28        0.0                  1.0         0.0    0.0
147          2024-07-29        3.0                  2.0         2.0    1.0
148          2024-07-30        2.0                 12.0         2.0    0.0
149          2024-07-31        2.0                  6.0         1.0    2.0

coffee_name  Cortado  Espresso  Hot Chocolate  Latte
0                0.0       0.0            3.0    2.0
1                0.0       0.0            0.0    1.0
2                2.0       0.0            2.0    2.0
3                0.0       1.0            0.0    2.0
4                1.0       0.0            4.0    3.0
..               …        …             …     …
145              0.0       2.0            0.0    2.0
146              0.0       1.0            0.0    1.0
147              0.0       0.0            2.0    1.0
148              3.0       2.0            0.0    3.0
149              4.0       0.0            0.0    7.0

[150 rows x 9 columns]
```

[87]:
```
daily_sales.iloc[:,1:].describe().T.loc[:,['min','max']]
```

[87]:
```
                     min   max
coffee_name
Americano            0.0   5.0
Americano with Milk  0.0  12.0
Cappuccino           0.0   9.0
Cocoa                0.0   2.0
Cortado              0.0   4.0
Espresso             0.0   4.0
```

```
Hot Chocolate        0.0   4.0
Latte                0.0   7.0
```

[89]:
```
hourly_sales =df.groupby(['hour']).count()['date'].reset_index().
↪rename(columns={'date':'count'})
```

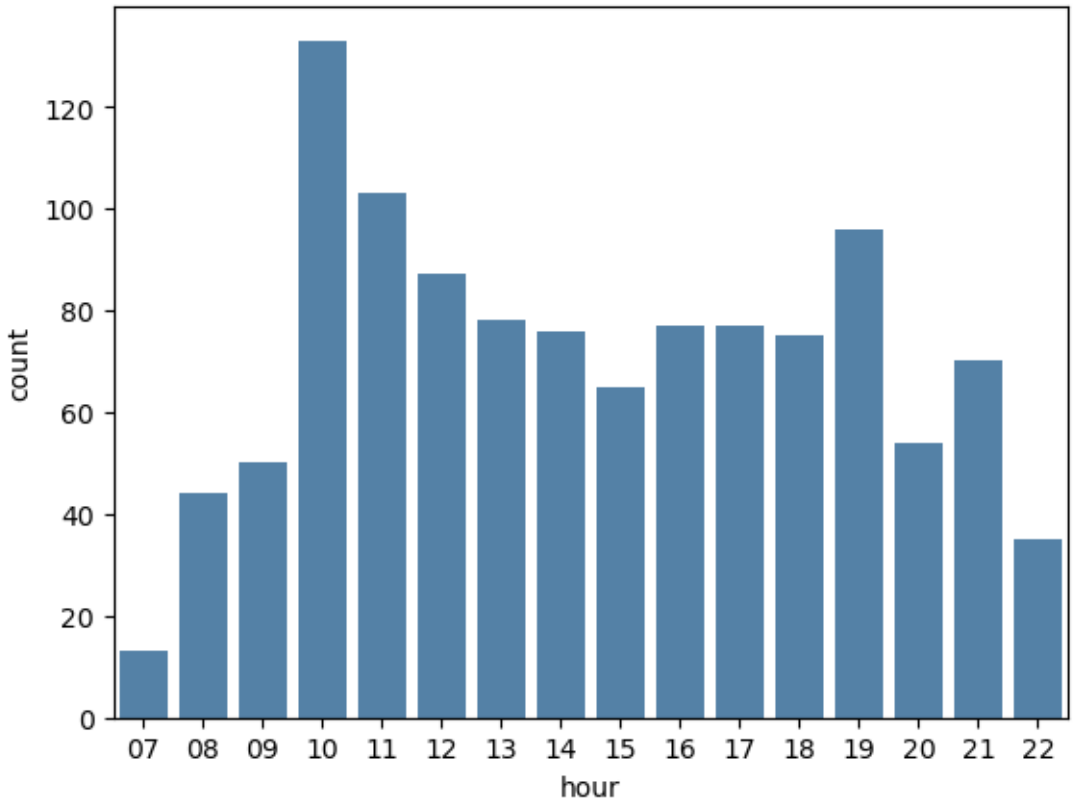
[91]:
```
hourly_sales
```

[91]:

	hour	count
0	07	13
1	08	44
2	09	50
3	10	133
4	11	103
5	12	87
6	13	78
7	14	76
8	15	65
9	16	77
10	17	77
11	18	75
12	19	96
13	20	54
14	21	70
15	22	35

[93]:
```
sns.barplot(data=hourly_sales,x='hour',y='count',color='steelblue')
```

[93]:
```
<Axes: xlabel='hour', ylabel='count'>
```



[95]:
```
hourly_sales_by_coffee =df.groupby(['hour','coffee_name']).count()['date'].
  ↪reset_index().rename(columns={'date':'count'}).
  ↪pivot(index='hour',columns='coffee_name',values='count').fillna(0).
  ↪reset_index()
```

[97]:
```
hourly_sales_by_coffee
```

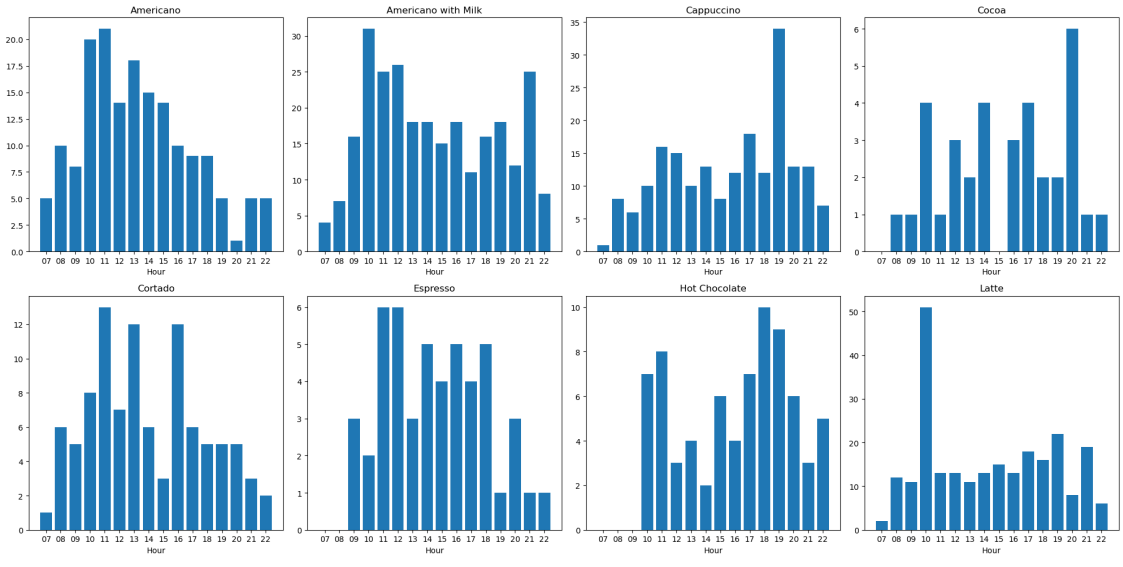
[97]:

coffee_name	hour	Americano	Americano with Milk	Cappuccino	Cocoa	Cortado	\
0	07	5.0	4.0	1.0	0.0	1.0	
1	08	10.0	7.0	8.0	1.0	6.0	
2	09	8.0	16.0	6.0	1.0	5.0	
3	10	20.0	31.0	10.0	4.0	8.0	
4	11	21.0	25.0	16.0	1.0	13.0	
5	12	14.0	26.0	15.0	3.0	7.0	
6	13	18.0	18.0	10.0	2.0	12.0	
7	14	15.0	18.0	13.0	4.0	6.0	
8	15	14.0	15.0	8.0	0.0	3.0	
9	16	10.0	18.0	12.0	3.0	12.0	
10	17	9.0	11.0	18.0	4.0	6.0	
11	18	9.0	16.0	12.0	2.0	5.0	
12	19	5.0	18.0	34.0	2.0	5.0	

```
13            20        1.0                 12.0        13.0    6.0      5.0
14            21        5.0                 25.0        13.0    1.0      3.0
15            22        5.0                  8.0         7.0    1.0      2.0
```

```
coffee_name  Espresso  Hot Chocolate  Latte
0                 0.0            0.0    2.0
1                 0.0            0.0   12.0
2                 3.0            0.0   11.0
3                 2.0            7.0   51.0
4                 6.0            8.0   13.0
5                 6.0            3.0   13.0
6                 3.0            4.0   11.0
7                 5.0            2.0   13.0
8                 4.0            6.0   15.0
9                 5.0            4.0   13.0
10                4.0            7.0   18.0
11                5.0           10.0   16.0
12                1.0            9.0   22.0
13                3.0            6.0    8.0
14                1.0            3.0   19.0
15                1.0            5.0    6.0
```

[103]:
```
fig, axs = plt.subplots(2, 4, figsize=(20, 10))
axs = axs.flatten()
for i, column in enumerate(hourly_sales_by_coffee.columns[1:]):
    axs[i].bar(hourly_sales_by_coffee['hour'],
hourly_sales_by_coffee[column])
    axs[i].set_title(f'{column}')
    axs[i].set_xlabel('Hour')
plt.tight_layout()
plt.show()
```



[ ]: