Problem Statement

Take out the keysights from this dataframe and make the prediction based on the data

Data Collection

The columns of following data are given below

```
date - shows the data of transaction
datetime - show datetime of transaction
cash-type - show how amount is paid cash or card
card - show the card number
money - show how much money card have
coffeename - shows the coffee name that is purchases
```

Techniques used

```
1 I have used pandas for data handling
2 I have used seaborn and matplotlib for exploratory data analysis
```

In [18]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
data = pd.read_csv('/content/excel index.csv')
print(data.head())

data.isnull().sum()

data = data.dropna()

data.isnull().sum()
```

```
         date                 datetime cash_type                 card  money  \
0  2024-03-01  2024-03-01 10:15:50.520      card  ANON-0000-0000-0001   38.7
1  2024-03-01  2024-03-01 12:19:22.539      card  ANON-0000-0000-0002   38.7
2  2024-03-01  2024-03-01 12:20:18.089      card  ANON-0000-0000-0002   38.7
3  2024-03-01  2024-03-01 13:46:33.006      card  ANON-0000-0000-0003   28.9
4  2024-03-01  2024-03-01 13:48:14.626      card  ANON-0000-0000-0004   38.7

     coffee_name
0          Latte
1  Hot Chocolate
2  Hot Chocolate
3      Americano
4          Latte
```

Out[18]:

	0
date	0
datetime	0
cash_type	0
card	0
money	0
coffee_name	0

dtype: int64

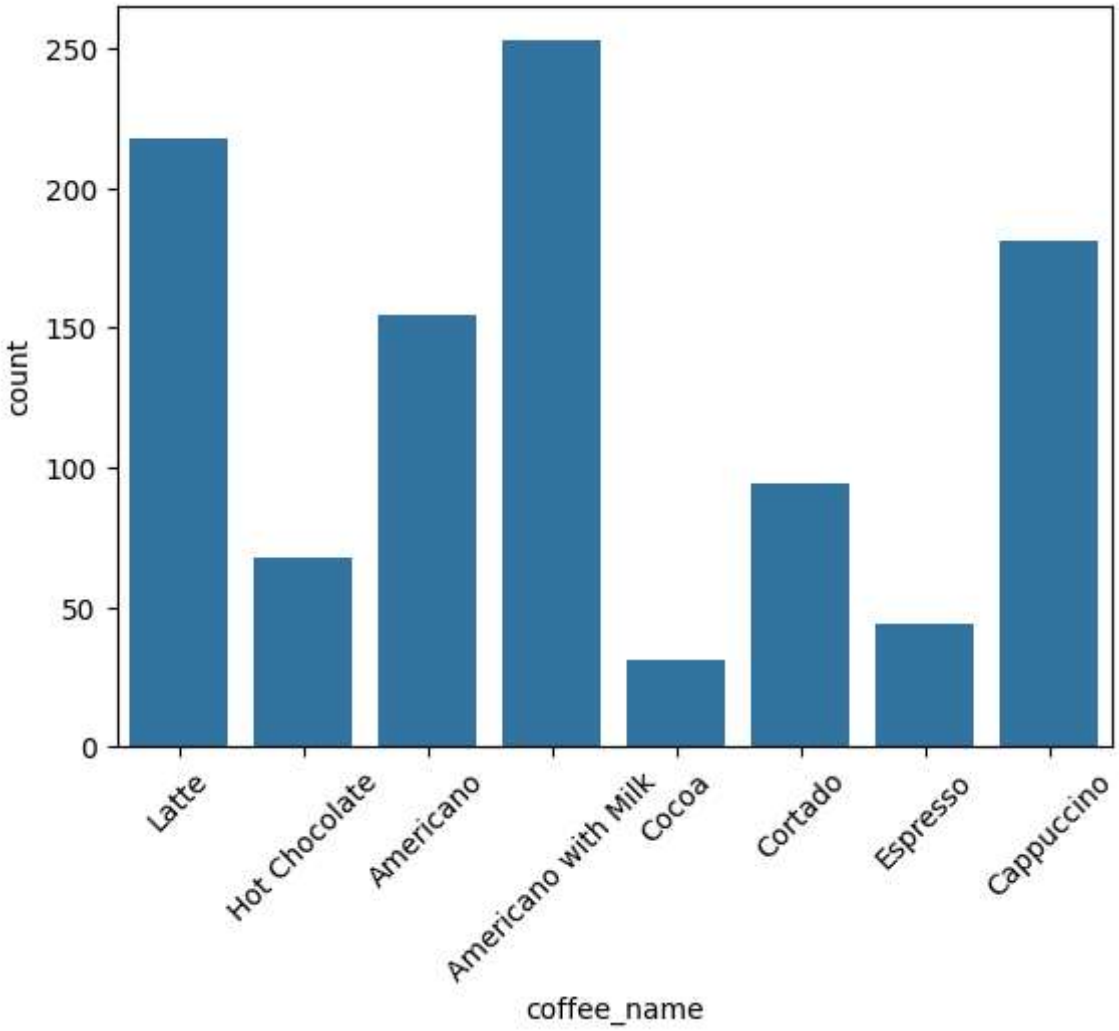
In [19]:
```
data.dtypes
```

Out[19]:

	0
date	object
datetime	object
cash_type	object
card	object
money	float64
coffee_name	object

dtype: object

In [20]:
```
countplot1 = sns.countplot(x = data['coffee_name'],data = data)
plt.xticks(rotation = 45)
plt.show()
```



Americano with Milk is the highly purchased product

New flavour should be added to Americano with Milk

Cocoa is the product with the lowest sales

so it is recommended to improve the marketing strategies of cocoa and improve the recipe of Cocoa and timely feedback of Cocoa coffee should be taken from the customers

It is also recommended to employ taster to improve the quality and sales of other coffee.

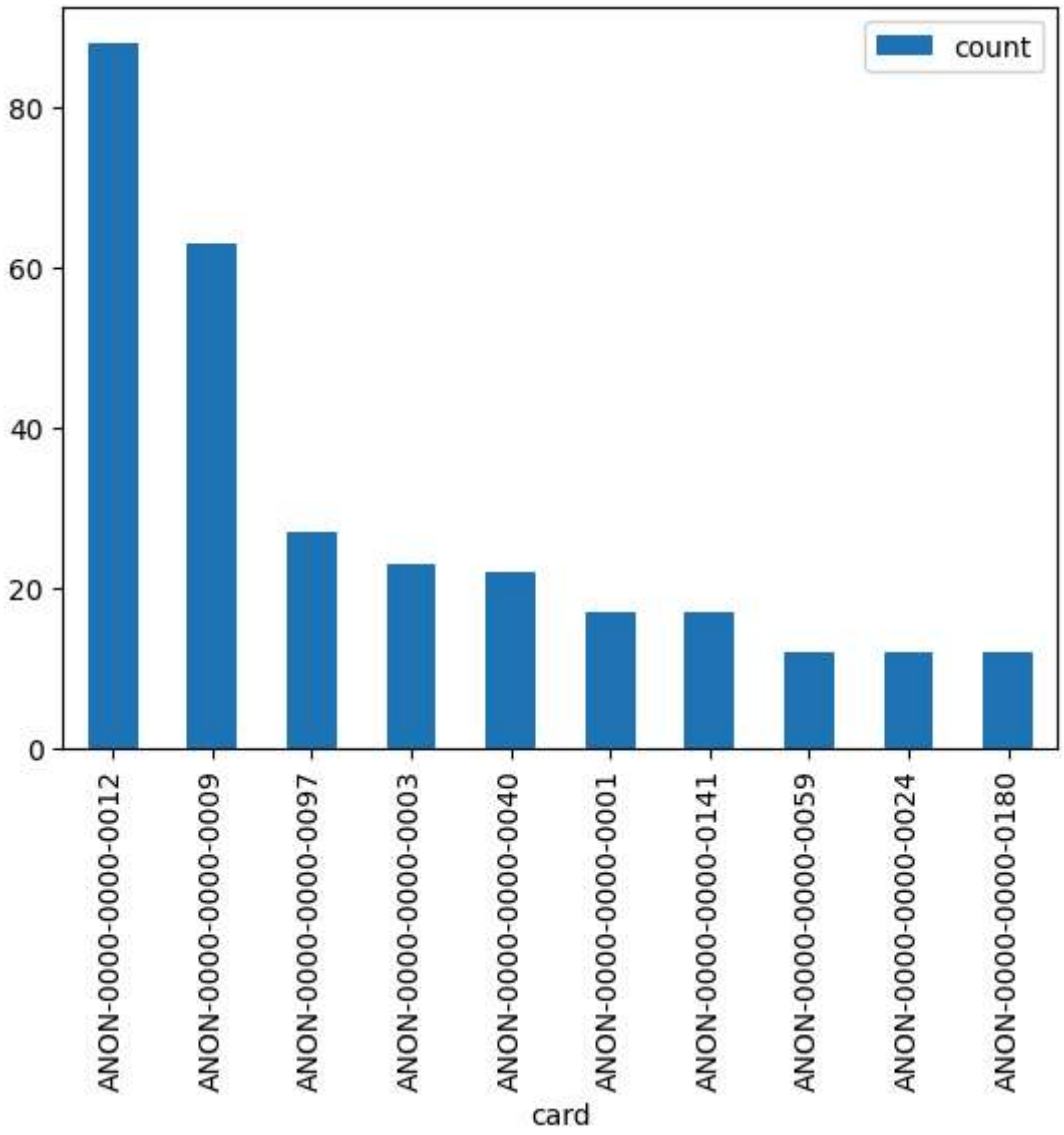
In [21]:
```
data['cash_type'].value_counts()
```

Out[21]:

cash_type	count
card	1044

dtype: int64

In [22]:
```
cardcount = pd.DataFrame(data['card'].value_counts().head(10))
cardcount.plot(kind = 'bar')
plt.show()
```



ANON-0000-0000-0012 is the card which has used highest amount to purchase coffee

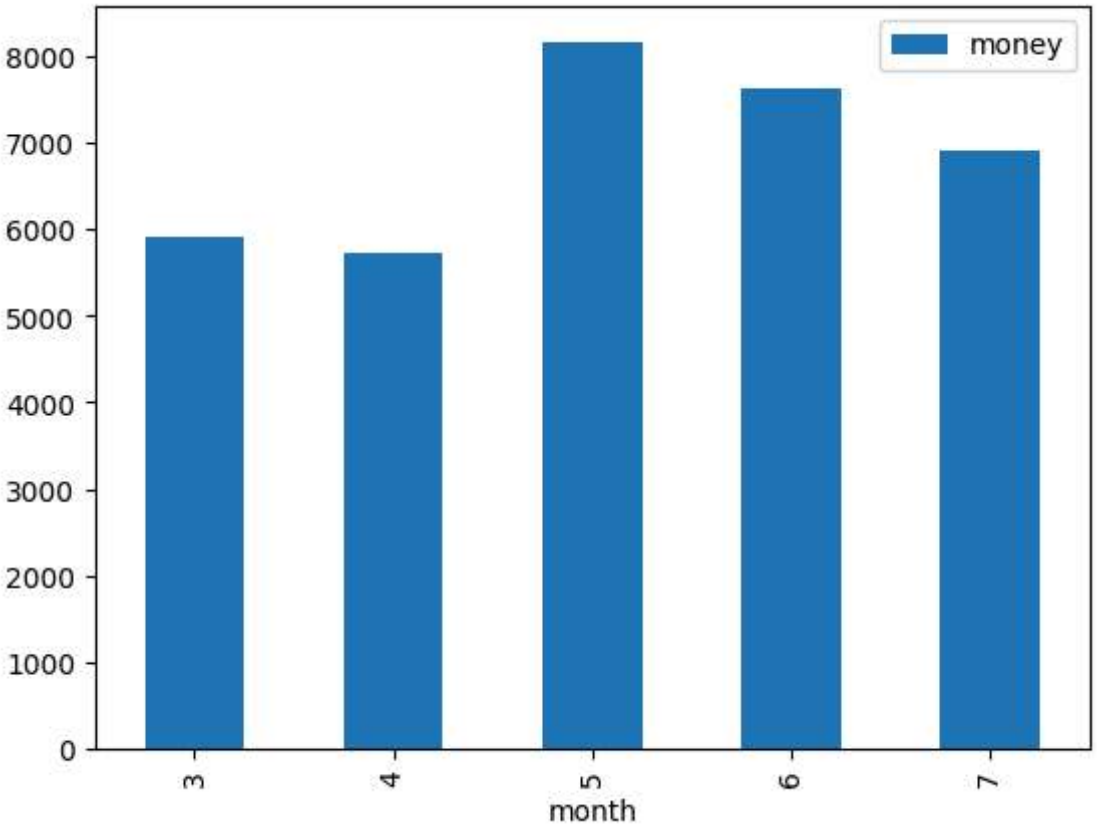
we can collaborate with bank which provide this type of card so that they can get discount on the money they spent on coffee

In [23]:
```
data.columns
```

Out[23]:
```
Index(['date', 'datetime', 'cash_type', 'card', 'money', 'coffee_name'], dtype='ob
ject')
```

In [24]:
```
data.dtypes
data['datetime'] = pd.to_datetime(data['datetime'])
data.dtypes
data['date'] = pd.to_datetime(data['date'])
data.dtypes
```

```
data['month'] = data['datetime'].dt.month
data['year'] = data['datetime'].dt.year
data['day'] = data['datetime'].dt.day
data['hour'] = data['datetime'].dt.hour
data['minute'] = data['datetime'].dt.minute
data['second'] = data['datetime'].dt.second
data.head()

mostprofitablemonth = pd.DataFrame(data.groupby('month')['money'].sum())
mostprofitablemonth.plot(kind = 'bar')
plt.show()
```

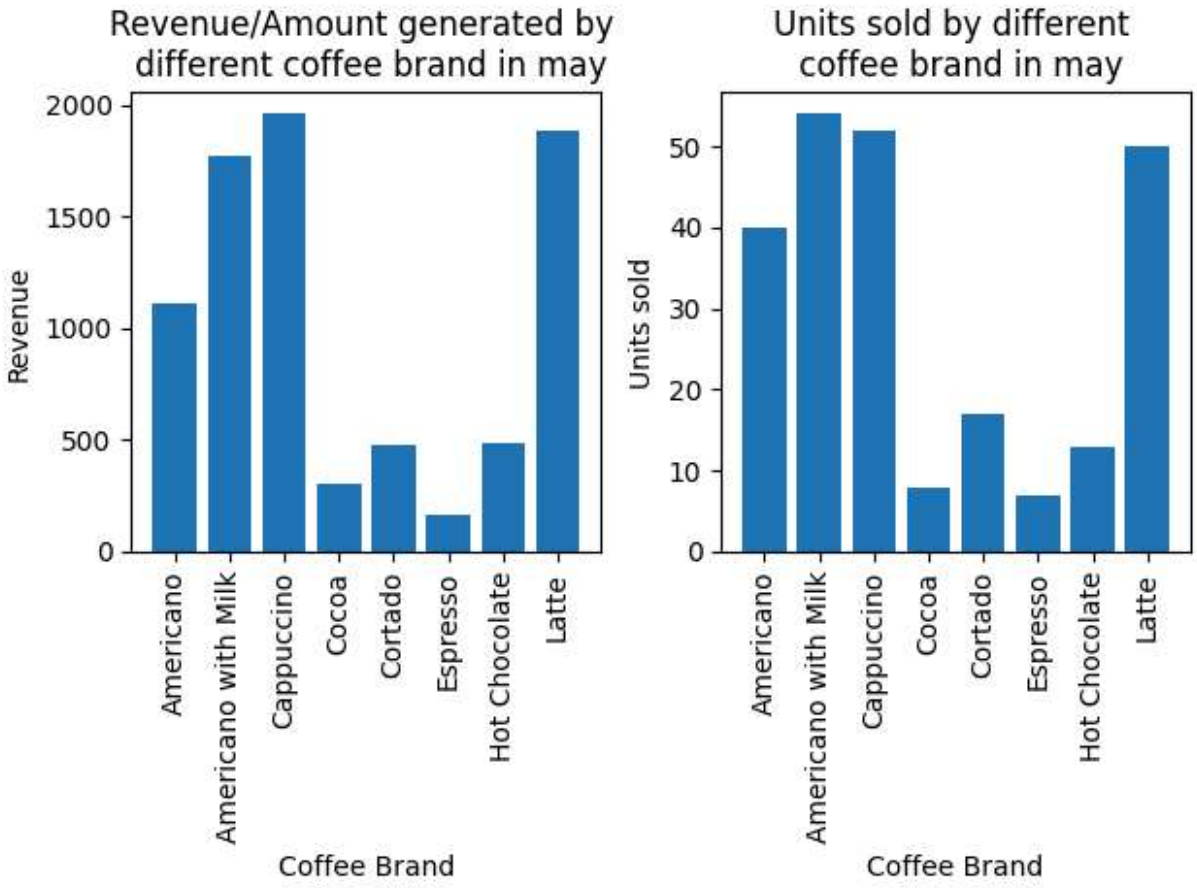


May is the most profitable month

So it recommended to may specific compaign should be launched such as offering discount on may on products and providing limited edition product on may.

In [25]:
```
maysalesproportion = pd.DataFrame(data[data['month'] == 5].groupby('coffee_name')['
plt.subplot(1,2,1)
plt.bar(maysalesproportion.index,maysalesproportion['money'])
plt.title("Revenue/Amount generated by \n different coffee brand in may")
plt.xlabel("Coffee Brand")
plt.ylabel("Revenue")
plt.xticks(rotation = 90)

mayunitsoldproportion = pd.DataFrame(data[data['month'] == 5].groupby('coffee_name'
plt.subplot(1,2,2)
plt.bar(mayunitsoldproportion.index,mayunitsoldproportion['coffee_name'])
plt.title("Units sold by different \n coffee brand in may")
```

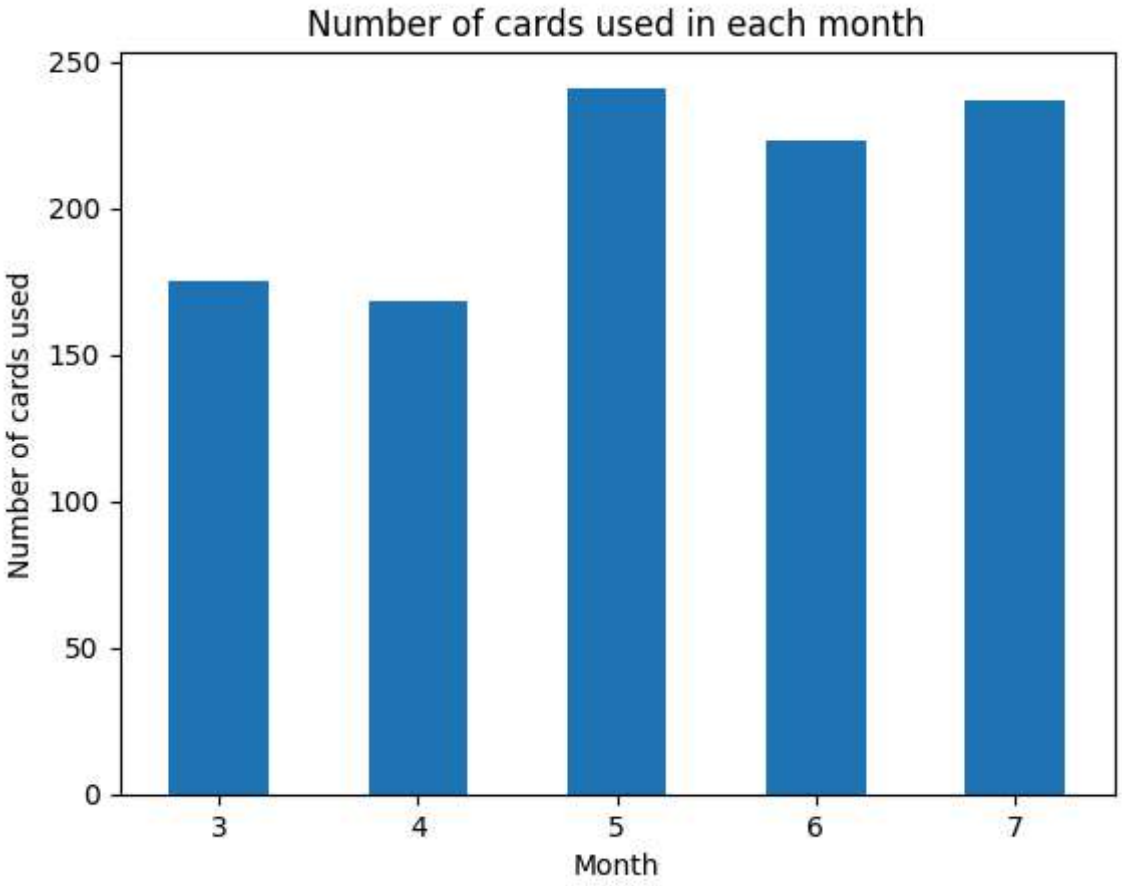
```
plt.xlabel("Coffee Brand")
plt.ylabel("Units sold")
plt.xticks(rotation = 90)
plt.tight_layout()
plt.subplots_adjust(hspace = 2)
plt.show()
```



In **May** the following insights are found

1. **Cappuccino** has generated the highest revenue in May
2. **Americano with Milk** is highly demanded by people in May

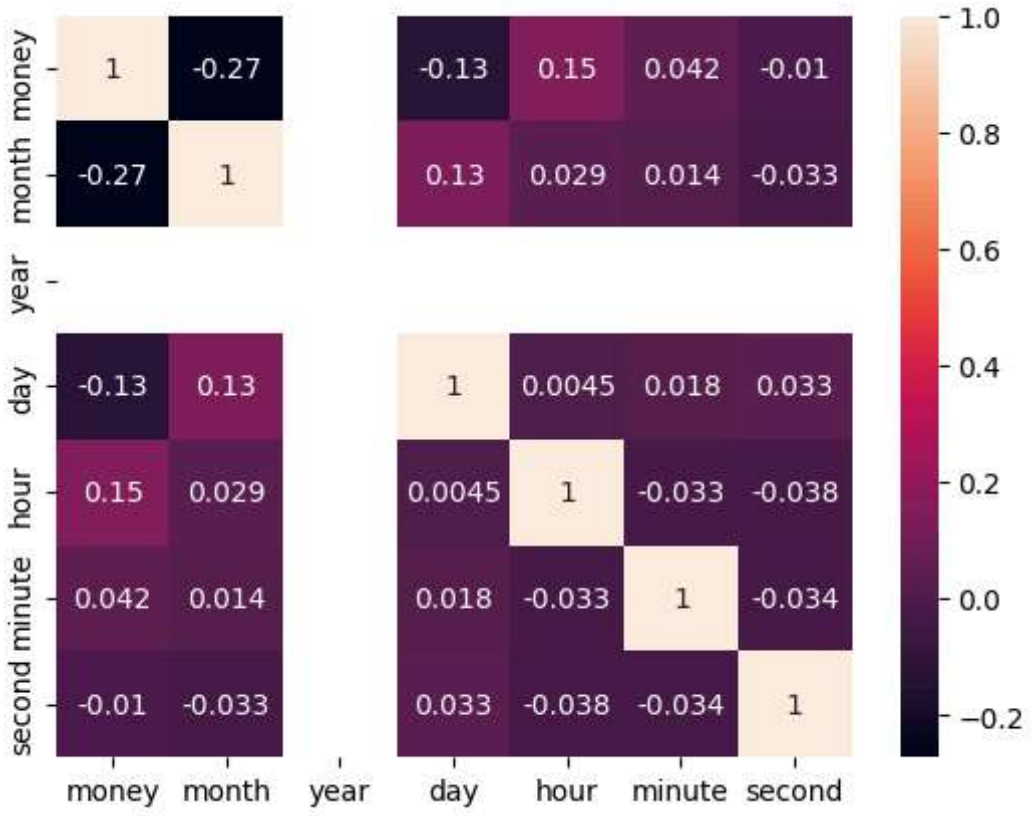
In [26]:
```
mostcardusingmonth = data.groupby('month')['card'].count()
plt.title("Number of cards used in each month")
mostcardusingmonth.plot(kind = 'bar')
plt.xlabel('Month')
plt.ylabel("Number of cards used")
plt.xticks(rotation = 0)
plt.show()
```



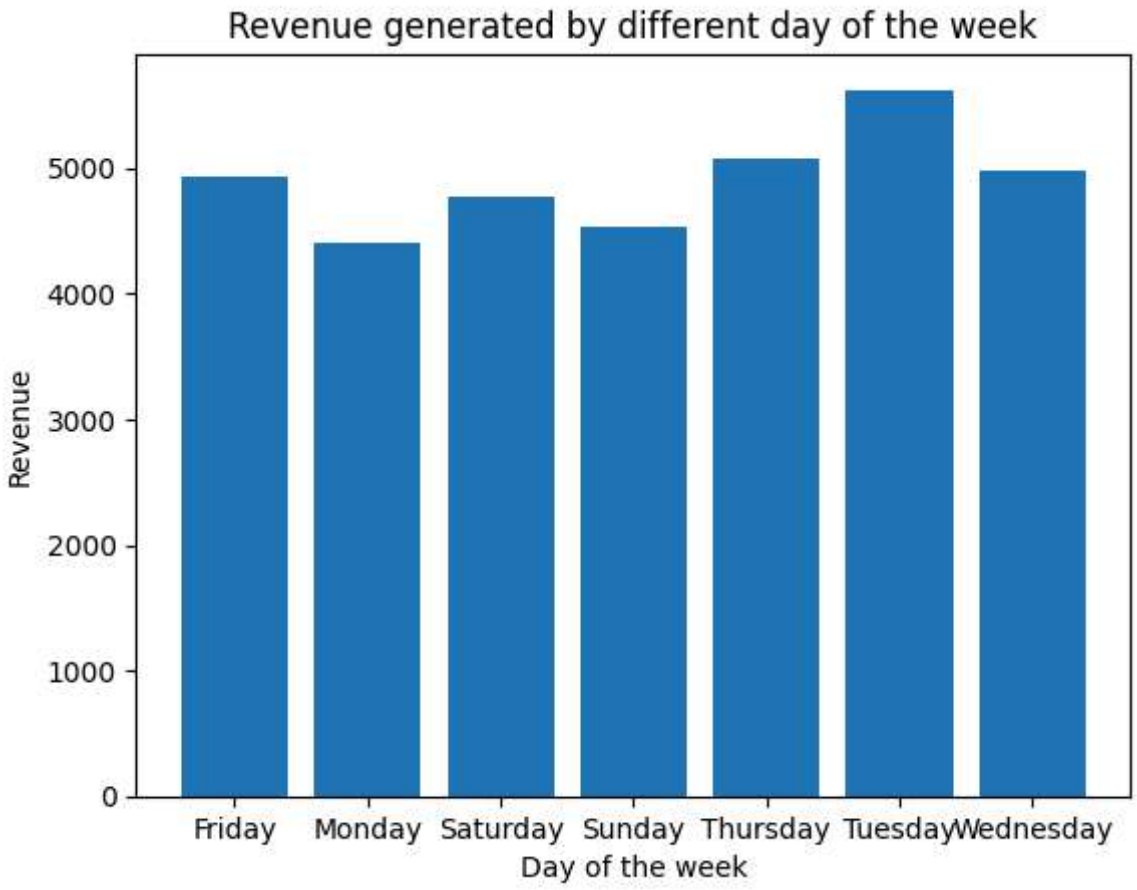
May is the month in which the most card are used for purchasing coffee

So it recommended to may specific compaign should be launched such as offering discount on may on products and providing limited edition product on may.

In [27]:
```
correlation = data.select_dtypes(include = [int,float]).corr()
sns.heatmap(correlation,annot = True)
plt.show()
```



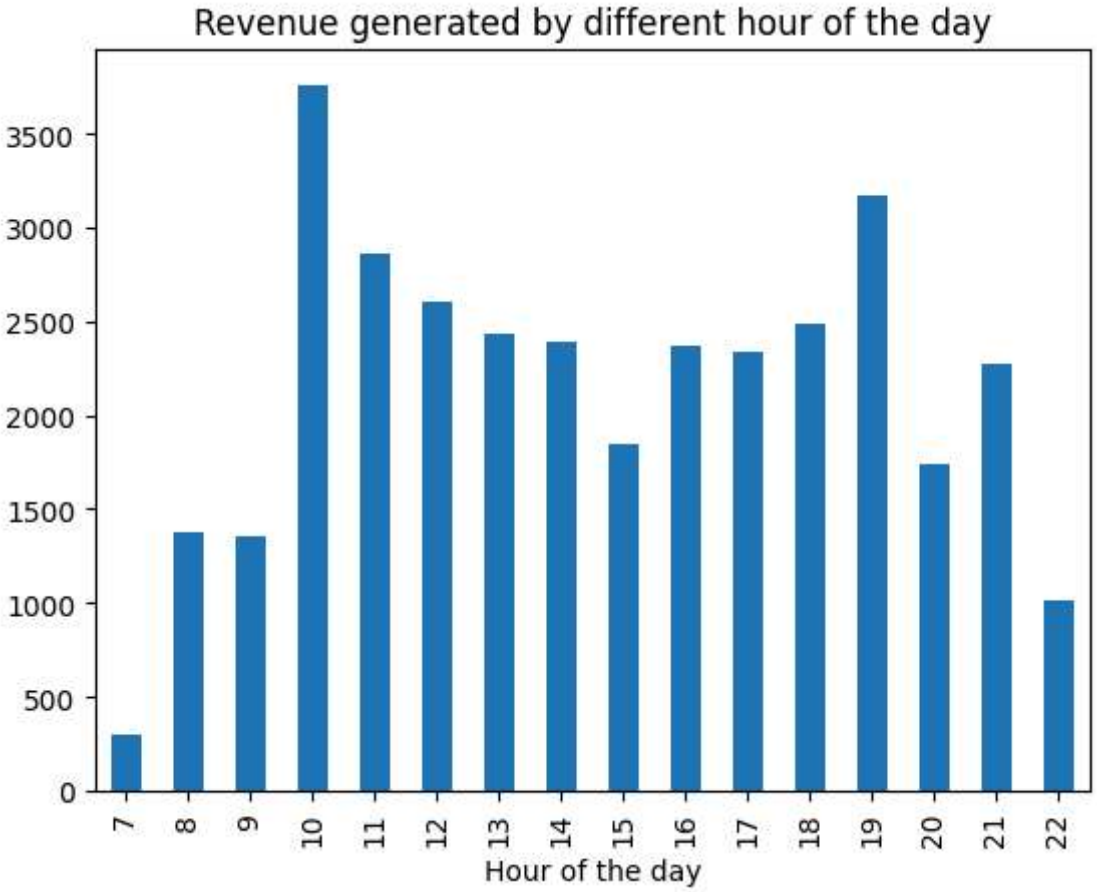
In [28]:
```
data['day_name'] = data['datetime'].dt.day_name()
data.head()

mostprofitableweekday = pd.DataFrame(data.groupby('day_name')['money'].sum().reset_
plt.bar(mostprofitableweekday['day_name'],mostprofitableweekday['money'])
plt.title("Revenue generated by different day of the week")
plt.xlabel("Day of the week")
plt.ylabel("Revenue")
plt.xticks(rotation = 0)
plt.show()
```



On Tuesday the sales of coffee is at the highest point and rest week day experience similar coffee sales

In [29]:
```
data.groupby('hour')['money'].sum().plot(kind = 'bar')
plt.title("Revenue generated by different hour of the day")
plt.xlabel("Hour of the day")
```

Out[29]:
```
Text(0.5, 0, 'Hour of the day')
```



Peak hours

1. From 10:00 to 11:00 the sales of coffee are the highest
2. From 07:00 to 08:00 the sales of coffee are the hightest

During the peak hour it is recommended to properly manage the staff and properly manage the inventory to prevent stock out and overcrowding of people.

In [29]: