Homework Session 15 – Data Exploratory

1. Please choose one data set from kaggle https://www.kaggle.com/
2. Please do data exploratory such as :
a. Checking Introductory Details About Data
b. Statistical Insight
c. Data Cleaning
d. Data Visualization
3. Please submit your task in github

1. **Data about Indian Food**

	name	ingredients	diet	prep_time	cook_time	flavor_profile	course	state	region
0	Balu shahi	Maida flour, yogurt, oil, sugar	vegetarian	45	25	sweet	dessert	West Bengal	East
1	Boondi	Gram flour, ghee, sugar	vegetarian	80	30	sweet	dessert	Rajasthan	West
2	Gajar ka halwa	Carrots, milk, sugar, ghee, cashews, raisins	vegetarian	15	60	sweet	dessert	Punjab	North
3	Ghevar	Flour, ghee, kewra, milk, clarified butter, su...	vegetarian	15	30	sweet	dessert	Rajasthan	West
4	Gulab jamun	Milk powder, plain flour, baking powder, ghee,...	vegetarian	15	40	sweet	dessert	West Bengal	East
...	...	...	...	...	...	...	...	...	...
250	Til Pitha	Glutinous rice, black sesame seeds, gur	vegetarian	5	30	sweet	dessert	Assam	North East
251	Bebinca	Coconut milk, egg yolks, clarified butter, all...	vegetarian	20	60	sweet	dessert	Goa	West
252	Shufta	Cottage cheese, dry dates, dried rose petals, ...	vegetarian	-1	-1	sweet	dessert	Jammu & Kashmir	North
253	Mawa Bati	Milk powder, dry fruits, arrowroot powder, all...	vegetarian	20	45	sweet	dessert	Madhya Pradesh	Central
254	Pinaca	Brown rice, fennel seeds, grated coconut, blac...	vegetarian	-1	-1	sweet	dessert	Goa	West

2. Indian cuisine consists of a variety of regional and traditional cuisines native to the Indian subcontinent. Given the diversity in soil, climate, culture, ethnic groups, and occupations, these cuisines vary substantially and use locally available spices, herbs, vegetables, and fruits. Indian food is also heavily influenced by religion, in particular Hinduism, cultural choices and traditions.

   Column Description
   name : name of the dish
   ingredients : main ingredients used
   diet : type of diet - either vegetarian or non vegetarian
   prep_time : preparation time
   cook_time : cooking time
   flavor_profile : flavor profile includes whether the dish is spicy, sweet, bitter, etc

course : course of meal - starter, main course, dessert, etc
state : state where the dish is famous or is originated
region : region where the state belongs
Presence of -1 in any of the columns indicates NaN value.

Data cleaning

In [81]:
```
#data clean.. remove rows with null value  (value -1)


for x in df.index:
  if df.loc[x, "prep_time"] < 0:
    df.drop(x, inplace = True)


for x in df.index:
  if df.loc[x, "flavor_profile"] == '-1':
    df.drop(x, inplace = True)

for x in df.index:
  if df.loc[x, "region"] == '-1':
    df.drop(x, inplace = True)
df
```

Out[81]:

	name	ingredients	diet	prep_time	cook_time	flavor_profile	course	state	region
0	Balu shahi	Maida flour, yogurt, oil, sugar	vegetarian	45	25	sweet	dessert	West Bengal	East
1	Boondi	Gram flour, ghee, sugar	vegetarian	80	30	sweet	dessert	Rajasthan	West
2	Gajar ka halwa	Carrots, milk, sugar, ghee, cashews, raisins	vegetarian	15	60	sweet	dessert	Punjab	North
3	Ghevar	Flour, ghee, kewra, milk, clarified butter, su...	vegetarian	15	30	sweet	dessert	Rajasthan	West
4	Gulab jamun	Milk powder, plain flour, baking powder, ghee,...	vegetarian	15	40	sweet	dessert	West Bengal	East
...	...	...	...	...	...	...	...	...	...
247	Prawn malai curry	Coconut milk, prawns, garlic, turmeric, sugar	non vegetarian	15	50	spicy	main course	West Bengal	East
249	Shukto	Green beans, bitter gourd, ridge gourd, banana...	vegetarian	10	20	spicy	main course	West Bengal	East
250	Til Pitha	Glutinous rice, black sesame seeds, gur	vegetarian	5	30	sweet	dessert	Assam	North East
251	Bebinca	Coconut milk, egg yolks, clarified butter, all...	vegetarian	20	60	sweet	dessert	Goa	West
253	Mawa Bati	Milk powder, dry fruits, arrowroot powder, all...	vegetarian	20	45	sweet	dessert	Madhya Pradesh	Central

190 rows × 9 columns

Statistic insight data

```
#find max cook time consumtion per region
df.groupby(['region'])['cook_time'].max().reset_index()
```

In [100]:
```
#find max cook time per region

df.groupby(['region'])['cook_time'].max().reset_index()
```

Out[100]:

	region	cook_time
0	Central	60
1	East	90
2	North	120
3	North East	45
4	South	120
5	West	720

```
#getting information how many various food in each region => west region has much various food

df['name'].groupby(df['region']).count().sort_values(ascending=False)
```

In [91]:
```
#getting information how many various food in each region => west region has much various food

df['name'].groupby(df['region']).count().sort_values(ascending=False)
```

Out[91]:
```
region
West          60
North         46
South         45
East          25
North East    10
Central        3
Name: name, dtype: int64
```

```
#count various taste => in india spicy flavor is favorite

df['name'].groupby(df['flavor_profile']).count().sort_values(ascending=False)
```

In [78]:
```
#count various taste => in india spicy flavor is favorite

df['name'].groupby(df['flavor_profile']).count().sort_values(ascending=False)
```

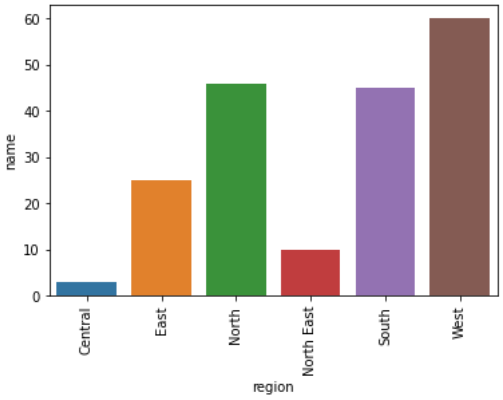
Out[78]:
```
flavor_profile
spicy     116
sweet      80
bitter      4
sour        1
Name: name, dtype: int64
```

Data visualitaion

```
#Total vaious food name in each region

df.groupby(['region'])['name'].count().reset_index()

df3 = df.groupby(['region'])['name'].count().reset_index()


sns.barplot(x=df3['region'], y=df3['name'])

plt.xticks(rotation="vertical")

plt.show()
```

In [111]:

```
#Total vaious food name in each region

df.groupby(['region'])['name'].count().reset_index()
df3 = df.groupby(['region'])['name'].count().reset_index()

sns.barplot(x=df3['region'], y=df3['name'])
plt.xticks(rotation="vertical")
plt.show()
```



In [105]:
```
#find max cook time per region

df.groupby(['region'])['cook_time'].max().reset_index()
```

Out[105]:

	region	cook_time
0	Central	60
1	East	90
2	North	120
3	North East	45
4	South	120
5	West	720

In [108]:
```
df2 = df.groupby(['region'])['cook_time'].max().reset_index()
df2['cook_time']  # create new dataframe for create new agregation



sns.barplot(x=df2['region'], y=df2['cook_time'])
plt.xticks(rotation="vertical")
plt.show()
```