

1. Data Analysis

```
In [1]: import pandas as pd
import numpy as np
```

```
In [2]: df=pd.read_csv('IndianFoodDatasetCSV.csv')
```

```
In [3]: df.head()
```

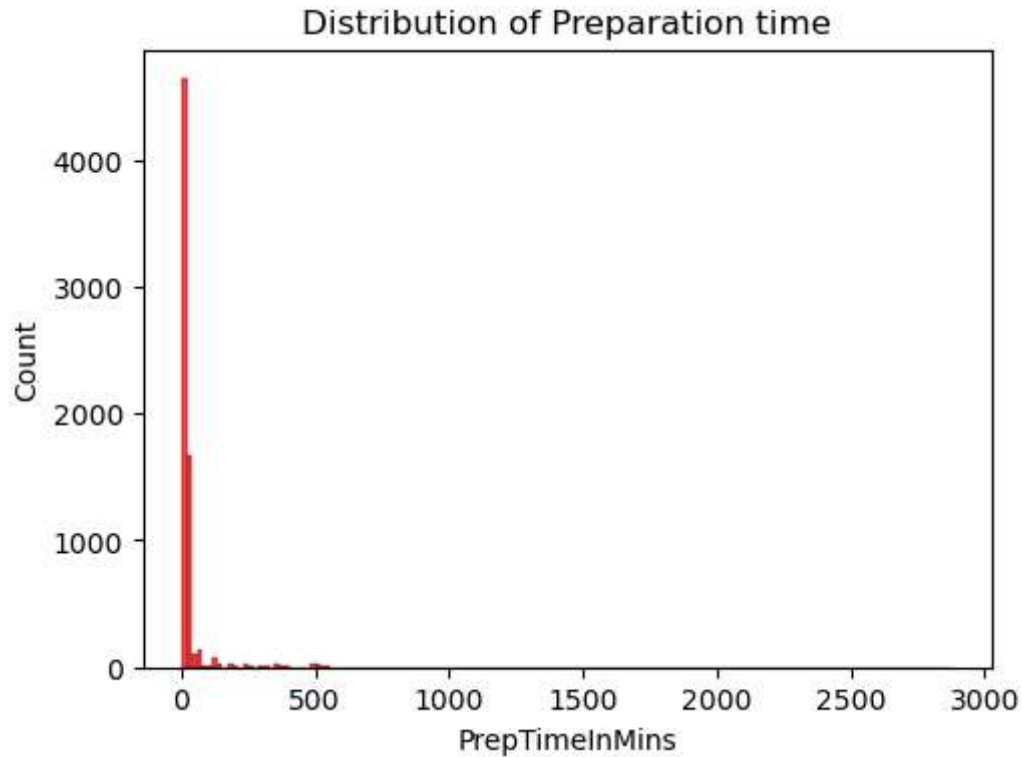
Out[3]:

	Srno	RecipeName	TranslatedRecipeName	Ingredients	TranslatedIngredients	PrepTimeInMins
0	1	Masala Karela Recipe	Masala Karela Recipe	6 Karela (Bitter Gourd/ Pavakkai) - deseeded,S...	6 Karela (Bitter Gourd/ Pavakkai) - deseeded,S...	15
1	2	टमाटर पुलियोगरे रेसिपी - Spicy Tomato Rice (Recipe)	Spicy Tomato Rice (Recipe)	2-1/2 कप चावल - पका ले,3 टमाटर,3 छोटा चमच बी...	2-1 / 2 cups rice - cooked, 3 tomatoes, 3 teas...	5
2	3	Ragi Semiya Upma Recipe - Ragi Millet Vermicel...	Ragi Semiya Upma Recipe - Ragi Millet Vermicel...	1-1/2 cups Rice Vermicelli Noodles (Thin),1 On...	1-1/2 cups Rice Vermicelli Noodles (Thin),1 On...	20
3	4	Gongura Chicken Curry Recipe - Andhra Style Go...	Gongura Chicken Curry Recipe - Andhra Style Go...	500 grams Chicken,2 Onion - chopped,1 Tomato -...	500 grams Chicken,2 Onion - chopped,1 Tomato -...	15
4	5	आंध्रा स्टाइल आलम पचड़ी रेसिपी - Adrak Chutney ...	Andhra Style Alam Pachadi Recipe - Adrak Chutn...	1 बड़ा चमच चना दाल,1 बड़ा चमच सफ़ेद उरद दाल,2...	1 tablespoon chana dal, 1 tablespoon white ura...	10

```
In [4]: import matplotlib.pyplot as plt # Visualization
import seaborn as sns #Visualization
```

```
In [5]: f= plt.figure(figsize=(12,4))  
  
ax=f.add_subplot(121)  
sns.histplot(df['PrepTimeInMins'],bins=150,color='r',ax=ax)  
ax.set_title('Distribution of Preparation time')
```

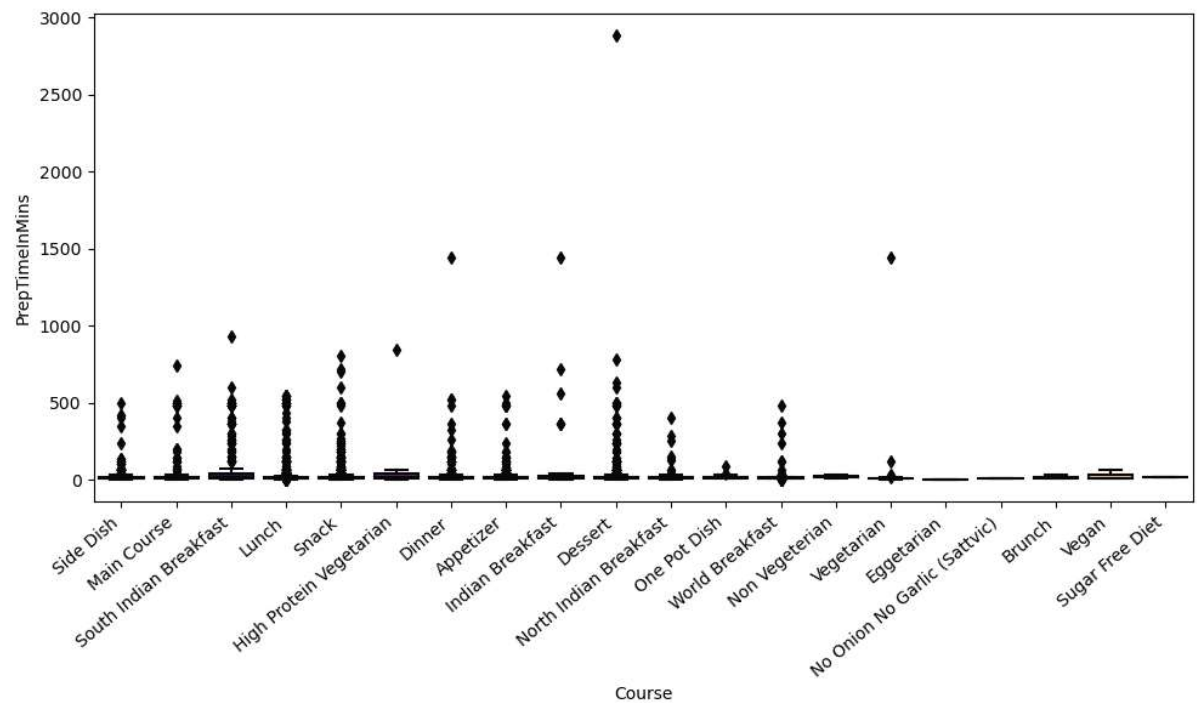
```
Out[5]: Text(0.5, 1.0, 'Distribution of Preparation time')
```



We can clearly notice from graph above that our data is highly biased.

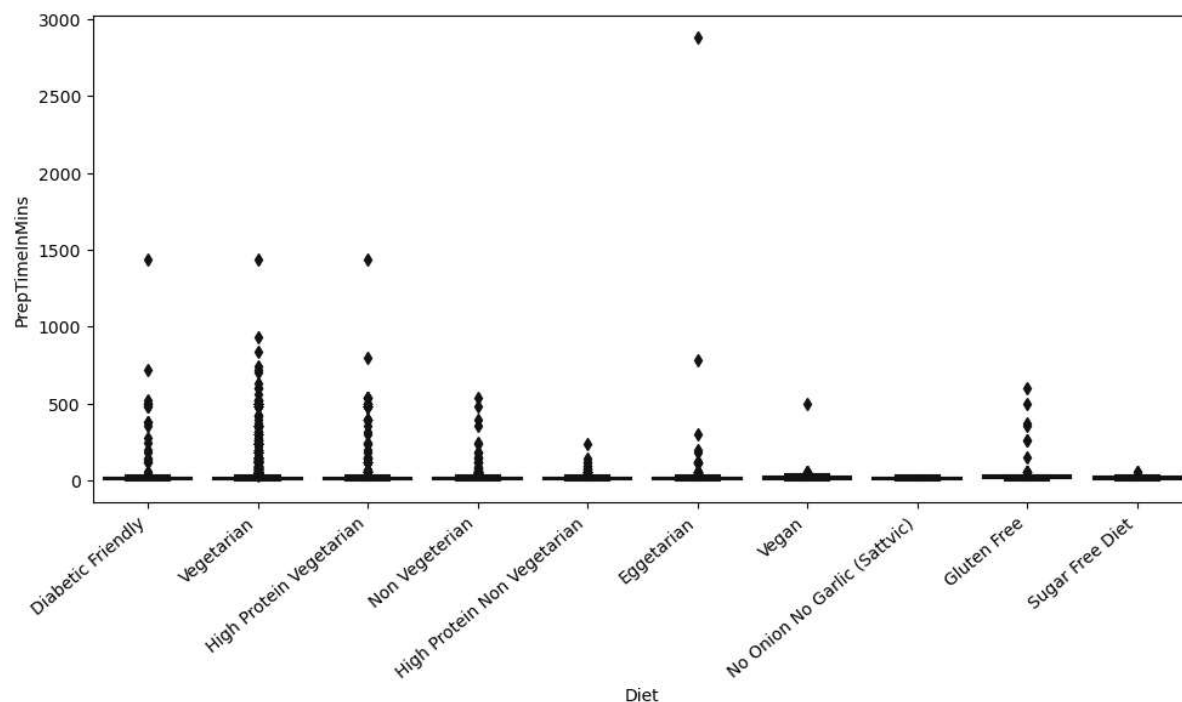
1.1 Lets visualize some graph with categorical values

```
In [9]: plt.figure(figsize=(10,6))
ax = sns.boxplot(x='Course', y='PrepTimeInMins',data=df,palette='magma')
ax.set_xticklabels(ax.get_xticklabels(), rotation=40, ha="right")
#sns.boxplot
plt.tight_layout()
plt.show()
```



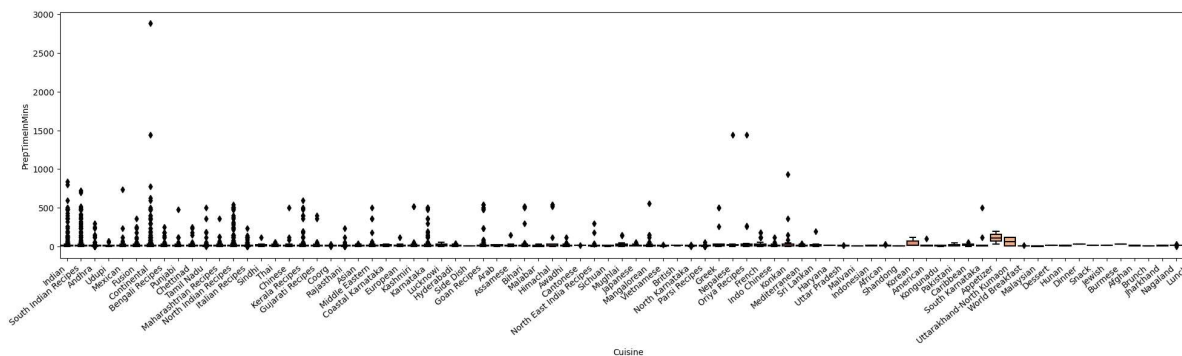
```
In [10]: plt.figure(figsize=(10,6))
ax = sns.boxplot(x='Diet', y='PrepTimeInMins',data=df,palette='magma')
ax.set_xticklabels(ax.get_xticklabels(), rotation=40, ha="right")

plt.tight_layout()
plt.show()
```



```
In [11]: plt.figure(figsize=(20,6))
ax = sns.boxplot(x='Cuisine', y='PrepTimeInMins',data=df,palette='magma')
ax.set_xticklabels(ax.get_xticklabels(), rotation=40, ha="right")

plt.tight_layout()
plt.show()
```

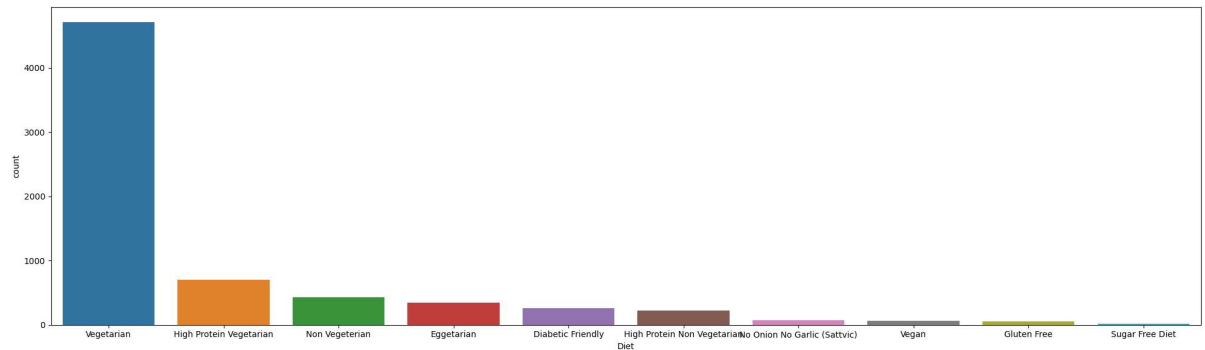


We learn following from the above graphs. 1) The data has many outliers (the diamond shaped points) 2) Most values are distributed close to zero which we can infer from the plot of 'Distribution of Preparation time'.

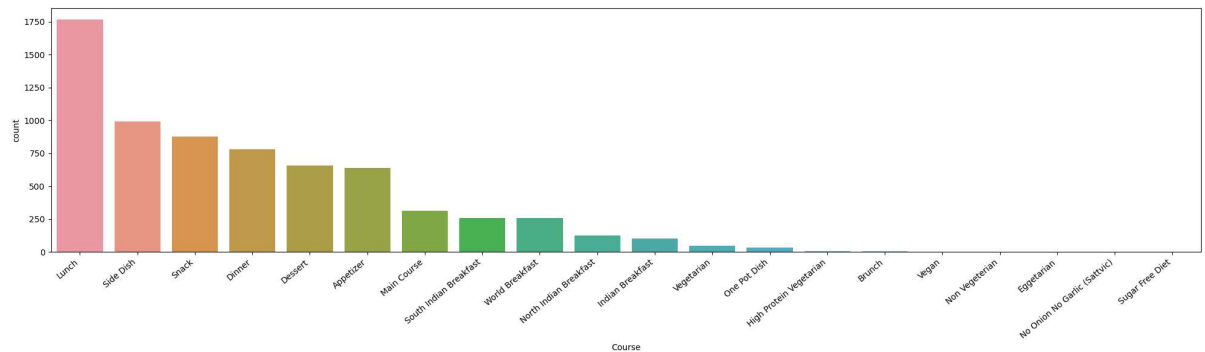
Getting some count plot on categorical parameters

```
In [12]: plt.figure(figsize=(20,6))
ax = sns.countplot(data=df, x="Diet" , order = df['Diet'].value_counts().index)

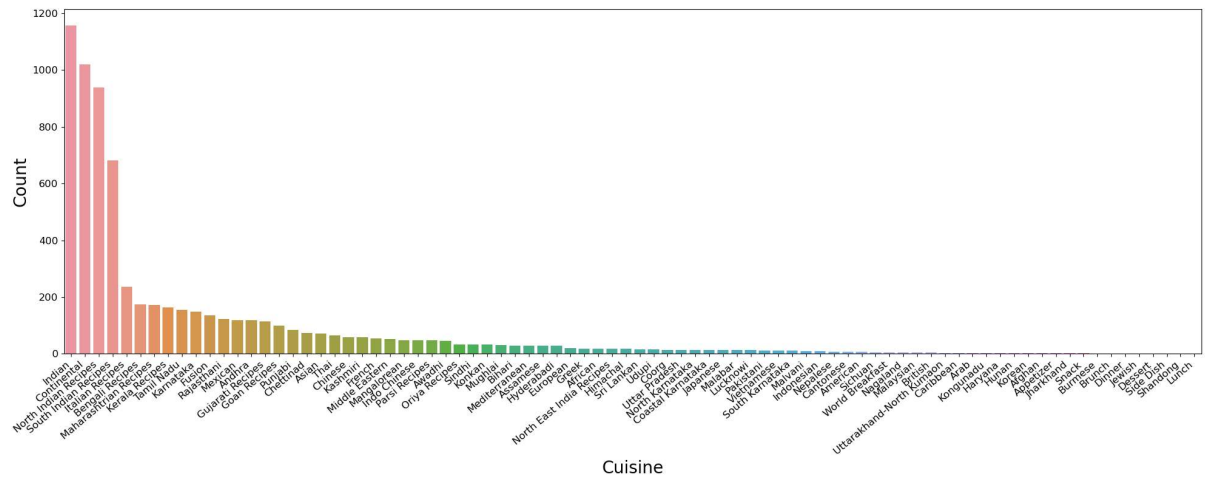
plt.tight_layout()
plt.show()
```



```
In [13]: plt.figure(figsize=(20,6))
ax = sns.countplot(data=df, x="Course" , order = df['Course'].value_counts().index)
ax.set_xticklabels(ax.get_xticklabels(), rotation=40, ha="right")
plt.tight_layout()
plt.show()
```



```
In [14]: plt.figure(figsize=(20,8))
ax = sns.countplot(data=df, x="Cuisine" , order = df['Cuisine'].value_counts
().index)
ax.set_xticklabels(ax.get_xticklabels(), rotation=40, ha="right")
ax.set_xlabel("Cuisine",fontsize=20)
ax.set_ylabel("Count",fontsize=20)
ax.tick_params(labelsize=12)
plt.tight_layout()
plt.show()
```



From the bar plots above, we can see that the most prevalent Diet is "Vegetarian", most prevalent Course is "Lunch", most prevalent Cuisine is "Indian".