Prediction of Visitors using Machine Learning Model

Aradhya Kanth

12/02/2023

Abstract

In this report, I have discussed how to effectively manage visitors, reduce wasteful expenditure and increase revenue.

The ability to make accurate forecasts on the number of customers is a pre-requisite for efficient planning and use of resources in various industries. It also contributes to global challenges of society such as food waste. Tourism is a domain particularly focussed on short-term forecasting for which the existing literature suggests that calendar and weather data are the most important sources for accurate prediction. We collected and make available a dataset with visitor counts over ten years from four different businesses representative for the tourism sector in Switzerland, along with nearly a thousand features comprising weather, calendar, event and lag information.

The leisure industry caters to millions of visitors every day, and each one of them arrives with their own set of expectations. Meeting those expectations is the key to getting people to return and, increasingly, leisure operators are turning to advanced analytics solutions for clues about how to keep their customers happy.

Problem Statement

Help leisure Industry such as Restaurants to predict visitor numbers accurately, and evaluate how these might be affected by weather, major sporting and public events, or other external factors. A better understanding of visitor numbers throughout the year can also be used for marketing purposes by targeting offers and promotions for the times they are needed most to boost numbers.

Market/Customer/Business Need Assessment

To increase revenue of a restaurant by predicting visitor count here are some steps to conduct a market, customer, and business needs assessment:

1. Market Assessment:
    - Analyse historical data to understand trends in visitor count, including seasonal patterns and peak times of day/week.

- Research local events, festivals, and holidays that may impact visitor traffic to your restaurant.

- Gather data on local competitors and their customer traffic patterns to identify opportunities to capture more visitors.

2. Customer Assessment:

    - Use historical data and customer feedback to develop customer personas, which can help you better understand your customers' needs and preferences.
    - Develop a survey or feedback mechanism to collect data on how weather affects customer behaviour, such as willingness to dine outdoors, menu choices, or timing of visits.
    - Use social media and other online platforms to gather real-time data on customer feedback and preferences related to weather conditions.

3. Business Assessment:

    - Use historical data and weather forecasts to create a predictive model for visitor traffic to your restaurant.

    - Analyse your restaurant's capacity and staffing need to ensure that you can meet increased demand during peak periods.
    - Develop a marketing plan to promote your restaurant during periods of favourable weather and leverage weather data to make data-driven decisions about staffing, menu offerings, and promotions.

Target Specification

To increase revenue of a restaurant by predicting visitor count and the effect of weather, it's important to develop a target specification that outlines your goals and the strategies you will use to achieve them.

1. Revenue Goals: Specify your revenue goals and the time frame for achieving them. For example, you may want to increase revenue by 10% in the next six months.

2. Visitor Count Goals: Specify your visitor count goals and the time frame for achieving them. For example, you may want to increase visitor count by 20% during the summer season.
3. Weather Metrics: Specify the weather metrics that will impact your restaurant's visitor count, such as temperature, precipitation, and wind. Identify the thresholds for these metrics that will trigger changes in your restaurant's operations, such as offering outdoor seating or adjusting the menu.
4. Marketing Strategies: Identify the marketing strategies that will help you achieve your revenue and visitor count goals, such as targeted advertising, social media promotions, or email marketing campaigns. Specify the budget and resources needed to implement these strategies.
5. Operations Strategies: Identify the operations strategies that will help you meet increased demand during peak periods, such as optimizing staffing levels, adjusting menu offerings, or expanding seating capacity. Specify the budget and resources needed to implement these strategies.
6. Data and Analytics: Specify the data and analytics tools you will use to measure progress towards your goals and make data-driven decisions about your marketing and operations strategies.

External Search

The sources I have used as reference for analysing the need of such a system, have mentioned below:

https://select-statistics.co.uk/business/leisure/

https://www.scitepress.org/Papers/2021/103230/103230.pdf

Benchmarking

In the market already many services are available such as eat app which forecast Restaurant sales but by their own way of calculating sales forecast. Generally, what it does is that it integrates POS with table management system like Eat App.

Applicable Patents

- Weather prediction method for forecasting selected events

The invention provides methods, systems, and computer program products for short term probability forecasting of selected weather-related events. These embodiments are adaptable for any geographical region that can be identified and for which a reasonable number of data points exist. The inventive method uses a data set of n observations of m parameters, where the parameters may be statistically correlated. https://patents.google.com/patent/US7069258B1/en

Applicable Regulations

- FSSAI guidelines for Restaurants & Eateries
  Food should be served with the appropriate equipment, and touching food with bare hands should be avoided. If food is maintained at room temperature, it should be consumed within 4 hours. Hot food that is served should be stored above 60-degree Celsius, and cold food should be kept below 5 degrees Celsius.

Applicable Constraints

- Data Availability: The accuracy and reliability of the revenue prediction model will depend on the quality and quantity of data available for analysis. Historical data on visitor count, revenue, and weather conditions will be required to train the model. The availability of data may also be limited by privacy concerns, legal regulations, or technical issues.
- Accuracy: The revenue prediction model needs to be highly accurate to be useful in practical scenarios. Any errors in the prediction can lead to significant financial losses for the restaurant, and a lack of accuracy may lead to the model not being used at all. The model should be validated using real-world data to ensure its accuracy.
- Scalability: The model should be able to scale up to handle large amounts of data and perform predictions in real-time. As the number of visitors increases, the model should be able to adapt to the changing environment and provide accurate predictions in a timely manner.
- Cost: The cost of developing and implementing the revenue prediction model should be reasonable and within the budget of the restaurant. This may involve finding cost-effective data sources, using open-source tools and platforms, and optimizing the model to reduce computational expenses.

- Ethical Considerations: Any data collected and analysed should be done so ethically, and with the consent of the visitors. Privacy concerns must be addressed, and the restaurant should take steps to ensure that the data is stored securely and not misused. Additionally, the model should be transparent, and the restaurant should be able to explain how the prediction is made.
- Weather Data Availability: The availability of weather data may also be a constraint, especially for small-scale restaurants that may not have access to real-time weather data. In such cases, the restaurant may need to rely on historical weather data or third-party sources for the weather data.

Business Model

They could use this model to effectively manage their staffs. For e.g. if considering all the phenomena my model predict large number of people might be coming today, the manager of the restaurant could hire some more staff that day and if less number of people from normal days then restaurant manager effectively manage it. Moreover, it will also help in food management.

They could allow local business to promote their product accordingly.

- Problem and Value Proposition: The problem we are addressing is the uncertainty around restaurant sales, which can be impacted by factors such as weather, holidays, and major sporting events. By using a machine learning model to predict sales based on these factors, restaurant owners can better plan staffing, inventory, and marketing efforts, and optimize their operations to increase sales.
- Customer Segments: The primary customer segments for this business model are restaurant owners and operators who are looking to optimize their operations and increase sales. This could include small independent restaurants as well as larger chains.
- Channels: The channels for delivering the machine learning model could include a web-based platform, mobile app, or integration with existing restaurant management systems. Marketing efforts could include online advertising, content marketing, and trade shows and conferences.

- Revenue Streams: The revenue streams for this business model could include subscription fees for access to the machine learning model and related tools, as well as revenue-sharing agreements with restaurant owners based on the increase in sales generated by the model.

Concept Generation

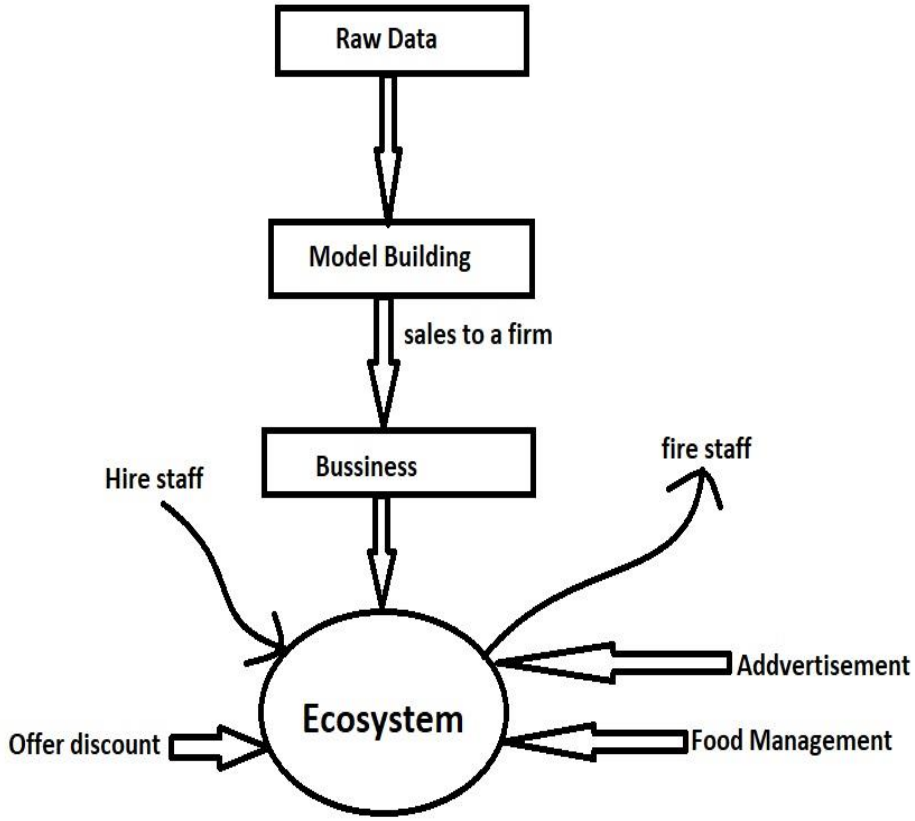
1. Data Collection: Collect historical data on restaurant sales, weather conditions (temperature, precipitation, etc.), holiday information, and major sporting events in the area.
2. Data Pre-processing: Clean the data, handle missing values, and convert categorical data into numerical data.
3. Feature Engineering: Extract features from the collected data that can help predict restaurant sales. For example, we can extract features such as day of the week, time of the day, weather conditions, and whether it is a holiday or a major sporting event.
4. Model Selection: Select an appropriate machine learning model for predicting restaurant sales. Some popular models include linear regression, decision trees, and neural networks.
5. Model Training: Train the selected model using the pre-processed data.
6. Model Evaluation: Evaluate the performance of the trained model using metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared.
7. Model Deployment: Deploy the trained model to make predictions on new data.

Concept Development

Build a ML model on a dataset containing various features such as public holiday, longitude, latitude, weather etc. and apply machine learning model to predict how these feature affect visitors coming to Restaurant.

Then give this model to Restaurant manager and they will use it to effectively run the Restaurant.

Final Product Prototype



Product details

Model building is the first work and then model is given to business firm.

It will use it to effectively manage their staff, food management, offer making, advertisement.

Data Source: From the past data

Algorithm: XGBRegressor

Code Implementation/Validation on Small Scale

Untitled - Jupyter Notebook

localhost:8888/notebooks/Feynn%20Lab/Project%201/Untitled.ipynb?kernel_name=python3#Visualisation

jupyter Untitled Last Checkpoint: an hour ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

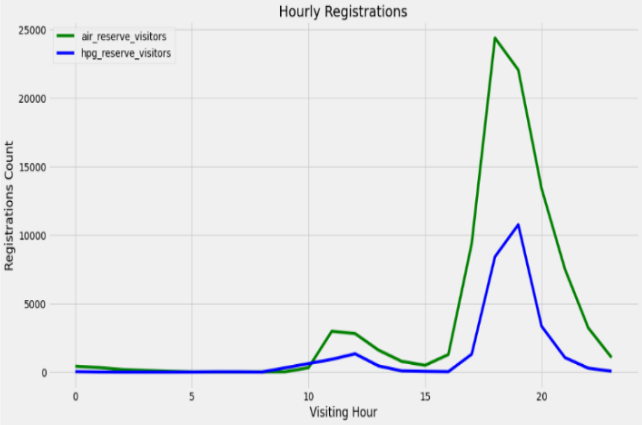
Trusted | Python 3 (ipykernel)

Run Code

In [89]:
```
# Hourly Visitors Behaviour
```

In [90]:
```
f, ax=plt.subplots(1,1, figsize=(15,8))
temp_1= air_reserve_data.groupby(['visit_hour'])['air_reserve_visitors'].count().to_frame()
temp_1.plot(kind='line',color= 'g', grid=True, ax=ax,legend=True)
temp_2= hpg_reserve_data.groupby(['visit_hour'])['hpg_reserve_visitors'].count().to_frame()
temp_2.plot(kind='line',color= 'b', grid=True, ax=ax,legend=True)
plt.xlabel('Visiting Hour')
plt.ylabel('Registrations Count')
plt.title('Hourly Registrations')
```

Out[90]: Text(0.5, 1.0, 'Hourly Registrations')



Feature Engineering

Untitled - Jupyter Notebook

localhost:8888/notebooks/Feynn%20Lab/Project%201/Untitled.ipynb?kernel_name=python3#Visualisation

Jupyter Untitled Last Checkpoint: an hour ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

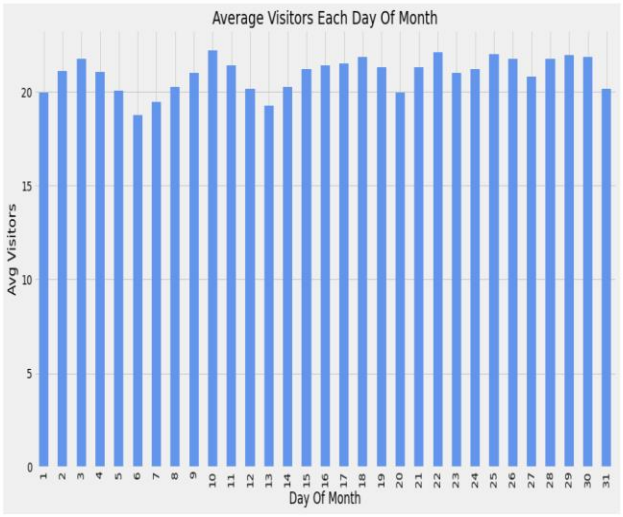
Trusted | Python 3 (ipykernel)

Run Code

In [87]:
```
# Average Visitors Each Day Of Month
```

In [88]:
```
f, ax=plt.subplots(1,1, figsize=(15,8))
temp= air_data.groupby(['day'])['visitors'].mean().to_frame()
temp.plot(kind='bar',color= 'cornflowerblue', grid=True, ax=ax,legend=False)
plt.xlabel('Day Of Month')
plt.ylabel('Avg Visitors')
plt.title('Average Visitors Each Day Of Month')
```

Out[88]: Text(0.5, 1.0, 'Average Visitors Each Day Of Month')



In [89]:
```
# Hourly Visitors Behaviour
```

In [90]:
```
f, ax=plt.subplots(1,1, figsize=(15,8))
```

Untitled - Jupyter Notebook

localhost:8888/notebooks/Feynn%20Lab/Project%201/Untitled.ipynb?kernel_name=python3#Visualisation

Jupyter Untitled Last Checkpoint: an hour ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

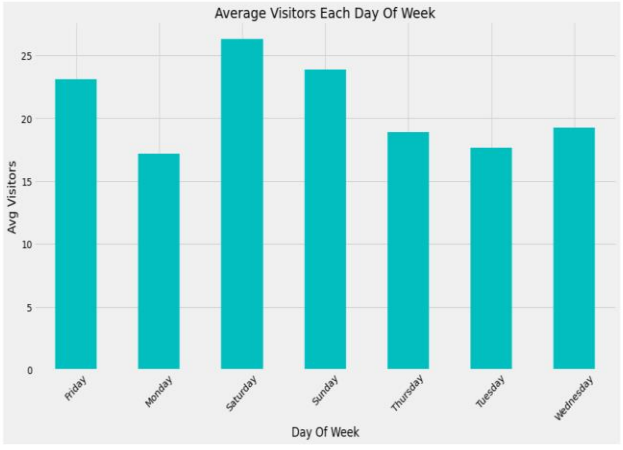
Trusted | Python 3 (ipykernel)

Run Code

In [84]:
```
# Average Visitors Each Day Of Week
```

In [85]:
```
f, ax=plt.subplots(1,1, figsize=(15,8))
temp= air_data.groupby(['day_of_week'])['visitors'].mean().to_frame()
temp.plot(kind='bar',color= 'c', grid=True, ax=ax, legend=False)
plt.xticks(rotation=45)
plt.xlabel('Day Of Week')
plt.ylabel('Avg Visitors')
plt.title('Average Visitors Each Day Of Week')
```

Out[85]: Text(0.5, 1.0, 'Average Visitors Each Day Of Week')



Untitled - Jupyter Notebook

localhost:8888/notebooks/Feynn%20Lab/Project%201/Untitled.ipynb?kernel_name=python3#Visualisation

jupyter Untitled Last Checkpoint: an hour ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3 (ipykernel)

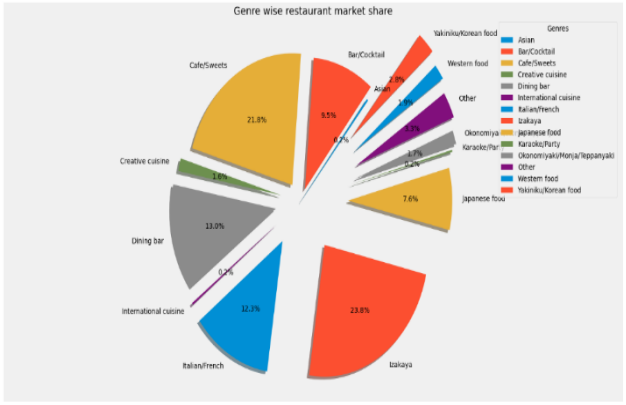
Run Code

In [82]:
```
temp = air_store_info.groupby(['air_genre_name'],as_index=False)['air_store_id'].count()
```

In [83]:
```
labels = temp.air_genre_name.values
sizes = temp.air_store_id.values

explode = (0, 0.1, 0.15, 0.2,0.25,0.3,0.35,0.4,0.45,0.5,0.55,0.6,0.65,0.7)
fig1, ax1 = plt.subplots(figsize=(20,10))
ax1.pie(sizes, explode=explode, labels=labels, autopct='%1.1f%%',
        shadow=True, startangle=50)

ax1.axis('equal')
plt.tight_layout()
plt.legend(title='Genres')
plt.title('Genre wise restaurant market share')
plt.show()
```



In [84]:
```
# Average Visitors Each Day Of Week
```

Untitled - Jupyter Notebook

localhost:8888/notebooks/Feynn%20Lab/Project%201/Untitled.ipynb?kernel_name=python3#Visualisation

jupyter Untitled Last Checkpoint: an hour ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

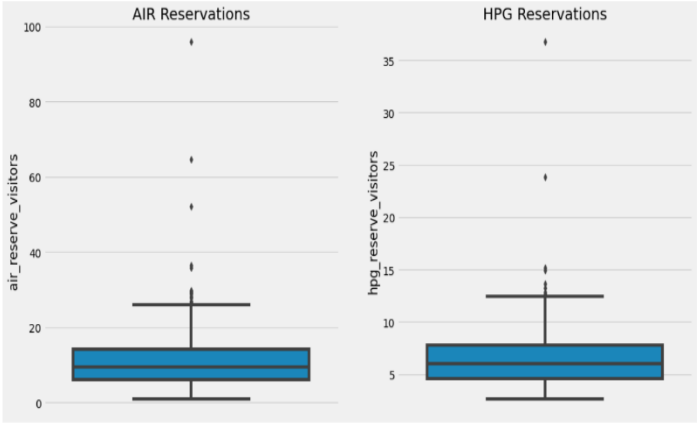
Trusted | Python 3 (ipykernel)

Run Code

Visitors Reserved

In [79]:
```
#  Boxplot of average visitors reservation per restaurant
```

In [80]:
```
f,ax = plt.subplots(1,2, figsize=(15,8))
sns.boxplot(y='air_reserve_visitors', data=temp_1,ax=ax[0])
plt.title("title")
sns.boxplot(y='hpg_reserve_visitors', data=temp_2,ax=ax[1])
ax[0].set_title('AIR Reservations')
ax[1].set_title('HPG Reservations')
plt.show()
```



Untitled - Jupyter Notebook

localhost:8888/notebooks/Feynn%20Lab/Project%201/Untitled.ipynb?kernel_name=python3#Visualisation

jupyter Untitled Last Checkpoint: an hour ago (autosaved)

Logout

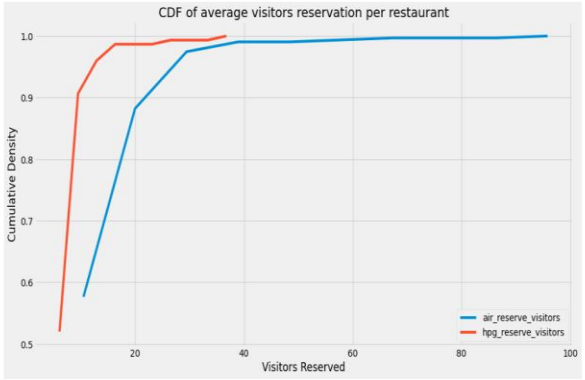
File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3 (ipykernel)

Run Code

In [78]:
```
f,ax = plt.subplots(1,1, figsize=(15,8))
counts, bin_edges = np.histogram(temp_1['air_reserve_visitors'], bins=10,
                                 density = True)
pdf = counts/(sum(counts))
cdf = np.cumsum(pdf)
plt.plot(bin_edges[1:], cdf,label='air_reserve_visitors')

counts, bin_edges = np.histogram(temp_2['hpg_reserve_visitors'], bins=10,
                                 density = True)
pdf = counts/(sum(counts))
cdf = np.cumsum(pdf)
plt.plot(bin_edges[1:], cdf,label='hpg_reserve_visitors')
plt.ylabel('Cumulative Density')
plt.xlabel('Visitors Reserved')
plt.legend()
plt.title('CDF of average visitors reservation per restaurant')
plt.show()
```



In [79]:
```
# Boxplot of average visitors reservation per restaurant
```

Untitled - Jupyter Notebook

localhost:8888/notebooks/Feynn%20Lab/Project%201/Untitled.ipynb?kernel_name=python3#Visualisation

jupyter Untitled Last Checkpoint: an hour ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

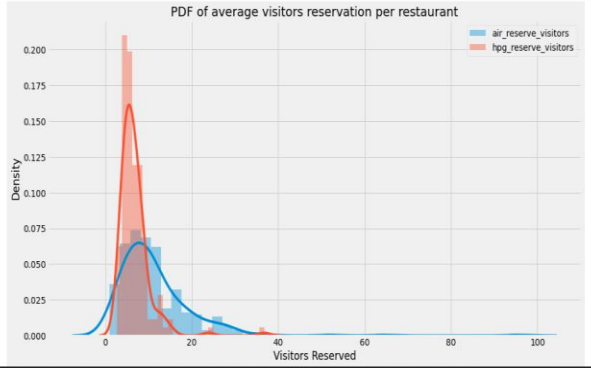
Trusted | Python 3 (ipykernel)

Run Code

In [77]:
```
temp_1 = total_air_reserve.groupby(['air_store_id'],as_index=False)['air_reserve_visitors'].mean()
temp_2 = total_hpg_reserve.groupby(['air_store_id'],as_index=False)['hpg_reserve_visitors'].mean()

f,ax = plt.subplots(1,1, figsize=(15,8))
sns.distplot(a=temp_1.air_reserve_visitors.values, ax=ax, label='air_reserve_visitors')
sns.distplot(a=temp_2.hpg_reserve_visitors.values, ax=ax, label='hpg_reserve_visitors')
plt.xlabel('Visitors Reserved')
plt.ylabel('Density')
plt.legend()
plt.title('PDF of average visitors reservation per restaurant')
plt.show()
```

```
C:\Users\aradh\anaconda3\lib\site-packages\seaborn\distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
C:\Users\aradh\anaconda3\lib\site-packages\seaborn\distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```



Untitled - Jupyter Notebook

localhost:8888/notebooks/Feynn%20Lab/Project%201/Untitled.ipynb?kernel_name=python3#Visualisation

jupyter Untitled Last Checkpoint: an hour ago (autosaved)

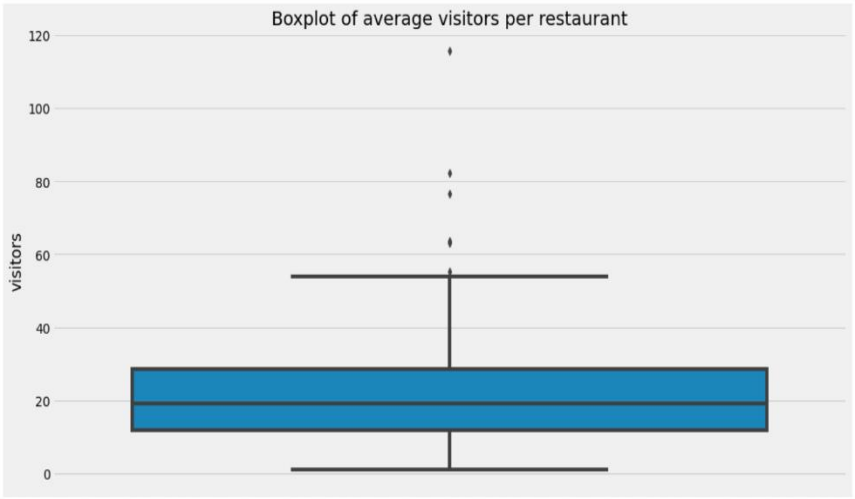
Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3 (ipykernel)

Run Code

In [76]:
```
f,ax = plt.subplots(1,1, figsize=(15,8))
sns.boxplot(y='visitors', data=temp,ax=ax)
plt.title('Boxplot of average visitors per restaurant')
plt.show()
```



Untitled - Jupyter Notebook

localhost:8888/notebooks/Feynn%20Lab/Project%201/Untitled.ipynb?kernel_name=python3#Visualisation

jupyter Untitled Last Checkpoint: an hour ago (autosaved)

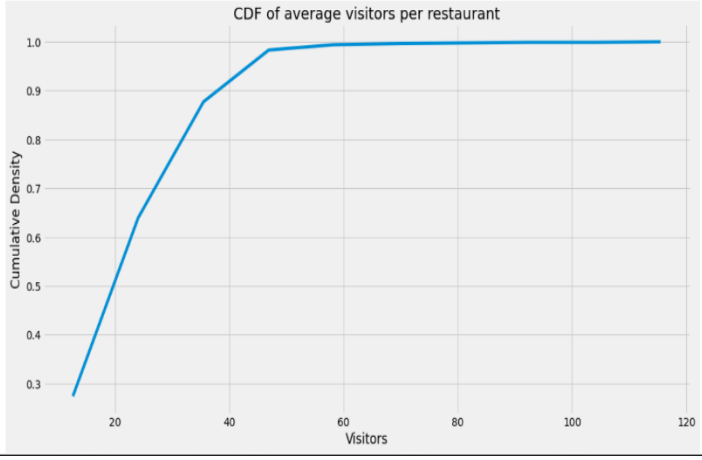
Logout

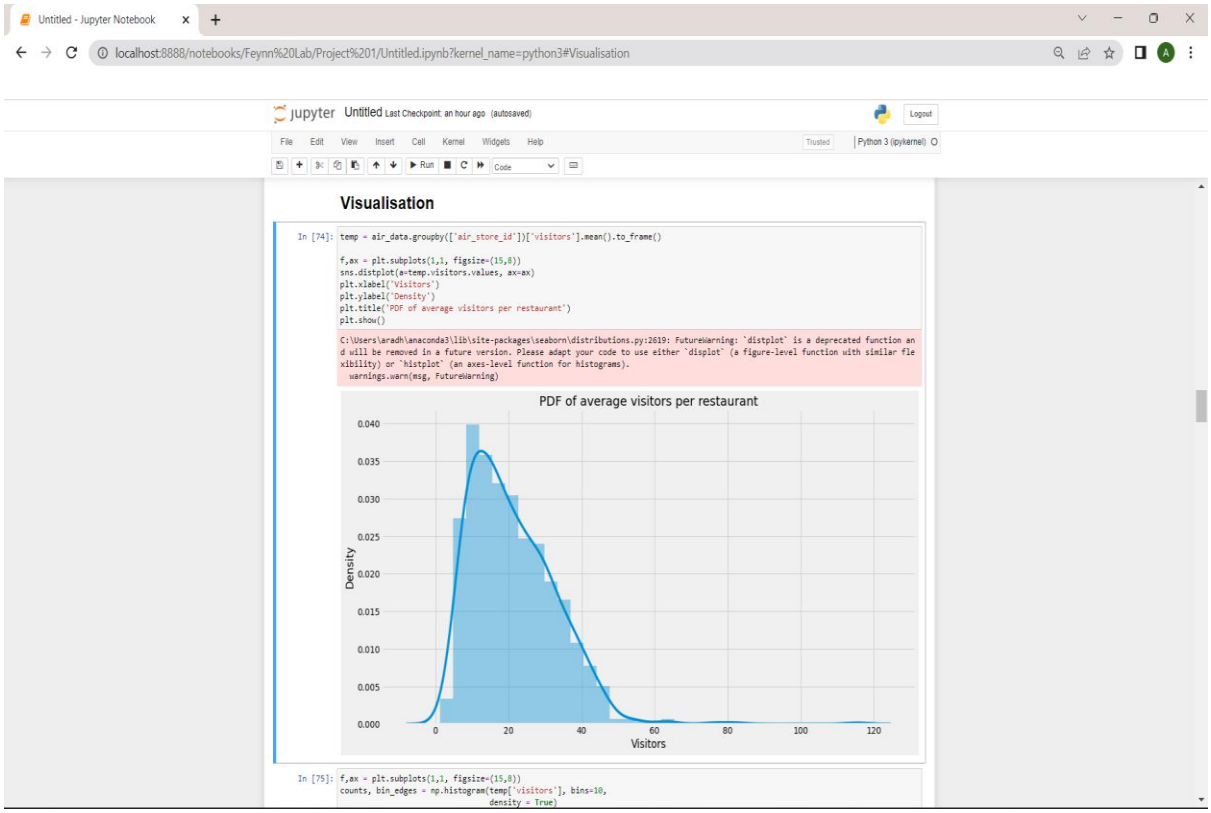
File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3 (ipykernel)

Run Code

In [75]:
```
f,ax = plt.subplots(1,1, figsize=(15,8))
counts, bin_edges = np.histogram(temp['visitors'], bins=10,
                                 density = True)
pdf = counts/(sum(counts))#col_pdf
cdf = np.cumsum(pdf)
plt.plot(bin_edges[1:], cdf)
plt.ylabel('Cumulative Density')
plt.xlabel('Visitors')
plt.title('CDF of average visitors per restaurant')
plt.show()
```





GitHub Link: https://github.com/aradhya-kanth/Feynn-Lab.git

Conclusion

Developing an ML model to predict the number of visitors to a restaurant can have a significant impact on the business. By accurately forecasting the number of customers, restaurant owners can better plan their resources, adjust staff schedules, and reduce waste.

ML model for predicting restaurant visitors can provide valuable insights and improve the efficiency and profitability of the business, making it a valuable tool for restaurant owners and managers.