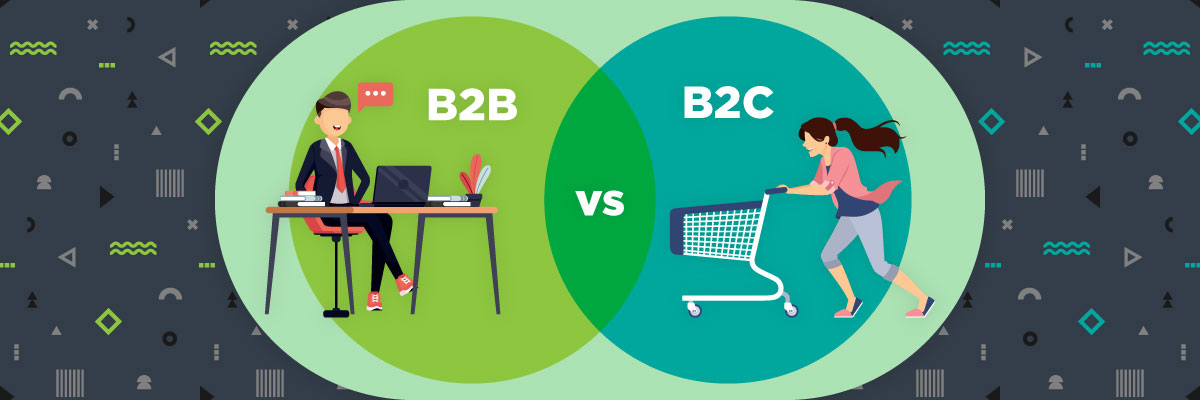
**PROJECT ON**

**EXPLORATORY DATA ANALYSIS ON AMAZON**

** **

**SUBMITTED BY:**

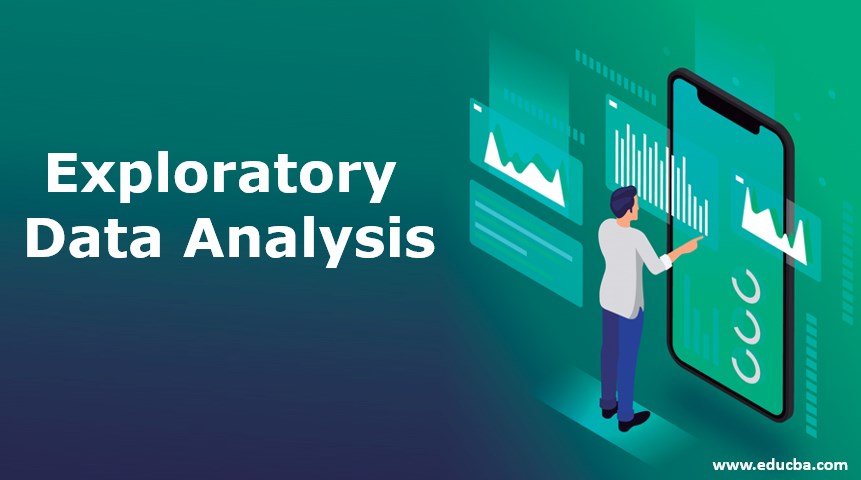
**NAME:** AKSHAYKUMAR.C

**EMP.ID:** EZ-1041

**DESIGNATION:** ASSOCIATE SOFTWARE ENGINEER

**MOB:**9611653780

**EXPLORATORY DATA ANALYSIS (EDA)**



**EDA:** IT IS A DATA ANALYTICS PROCESS TO UNDERSTAND THE DATA IN DEPTH AND LEARN THE DIFFERENT DATA CHARACTERISTICS ,OFTEN WITH VISUAL MEANS.

**THE FOUR MAJOR STEPS IN EDA :**

1. CLIENT REQUIREMENT CHECKING
2. DATA PRE-PROCESSING /DATA CLEANING
3. DATA VISUALIZATION
4. CONCLUSION FOR CLIENT

**DATA PRE-PROCESSING /DATA CLEANING**

**STEP 1:** IMPORTING LIBRARIES

**PANDAS:** IT IS PYTHON LIBRARY USED FOR WORKING WITH DATA SETS.

IT HAS FUNCTIONS FOR ANALYZING ,CLEANING ,EXPLORING,MANIPULATING DATA.

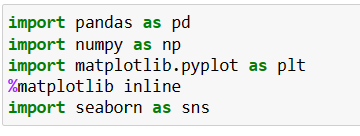
**NUMPY:** IT IS PYTHON LIBRARY USED FOR WORKING WITH ARRAYS.

NUMPY STANDS FOR NUMERICAL PYTHON.

**MATPLOTLIB:** IT IS A COMPREHENSIVE LIBRARY FOR CREATING STATIC ,ANIMATED ,INTERACTIVEVISUALIZATIONS IN PYTHON.

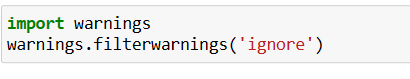
(**INLINE**: IT IS USED TO UPDATE IN THE SAME GRAPH /DATAFRAME)

**SEABORN**: IT IS A LIBRARY FOR MAKING STATISTICAL GRAPHICS IN PYTHON.



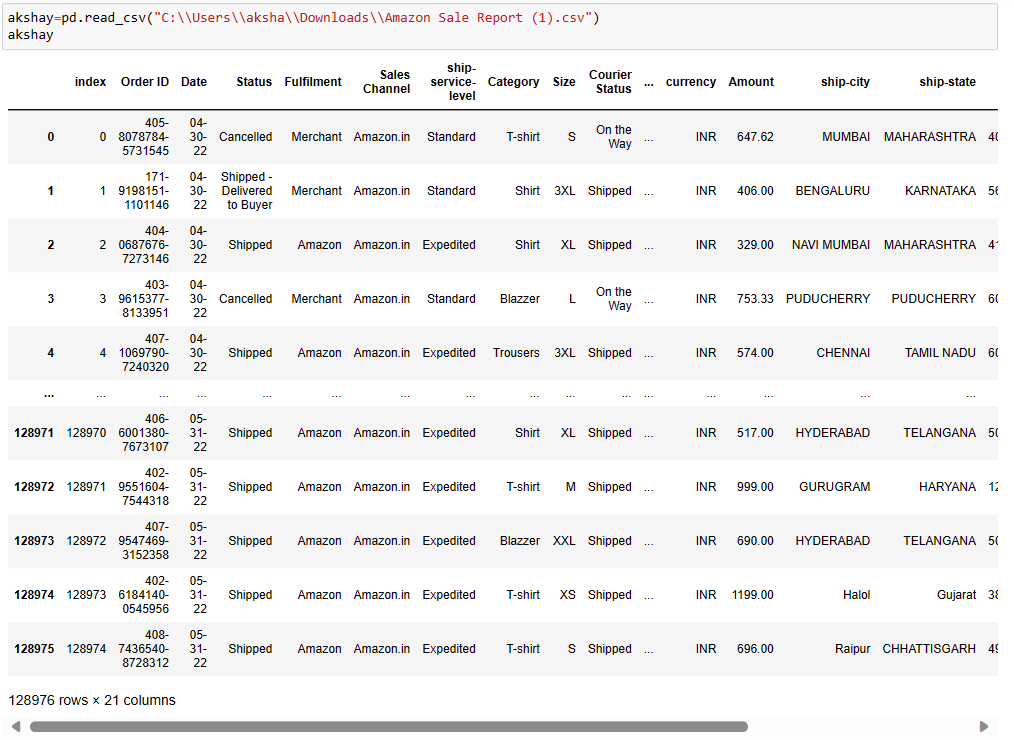
**STEP 2:** IMPORTING WARNINGS

WARNING MESSAGES ARE NORMALLY WRITTEN TO SYS.STDERR,BUT THEIR DISPOSITION CAN BE CHANGED FLEXIBLY, FROM IGNORING ALL WARNINGS TO TURNING THEM INTO EXCEPTIONS.



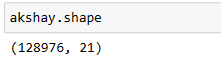
**STEP 3**: READING /IMPORTING DATA

USING PANDAS LIBRARY WE CAN IMPORT THE DATA , COPYING ITS PATH , BY CREATING / TAKING ANY VARIABLE NAME , STORING TO IT , AND THEN CALL THAT VARIABLE NAME .HERE WE MENTIONED PANDAS AS PD .



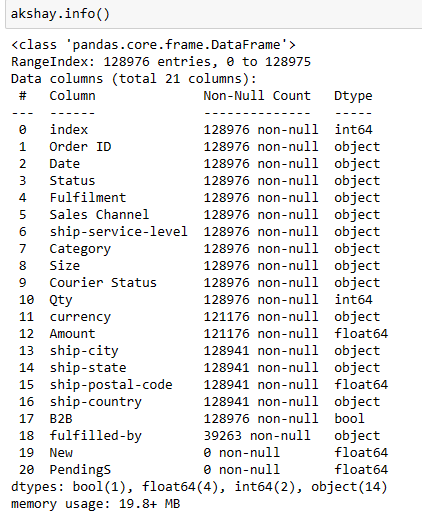
**STEP 4:** SHAPE OF THE DATAFRAME .

INTINAL SHAPE OF THE DATAFRAME CAN BE OBTAINED USING SHAPE , WE NEED TO DO THIS STEP BECAUSE AFTER REMOVING NULL /DUPLICATED VALUES …FOR CONFIRMATION PURPOSE ..WHELTHER THE NULL AND DUPLICATED VALUES ARE REMOVED .

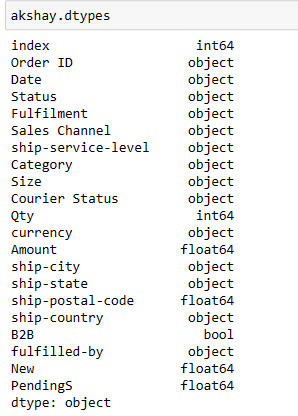


**STEP 5:** CHECKING THE INFORMATION OF THE DATAFRAME .

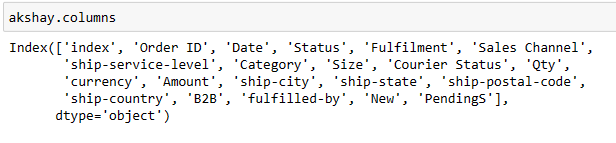
BASICALLY WE SHOULD CHECK THE INFORMATION OF THE CLIENT DATA … BY USING DATAFRAME NAME . INFO()…… WE CAN GET /FETCH THE RANGE INDEX



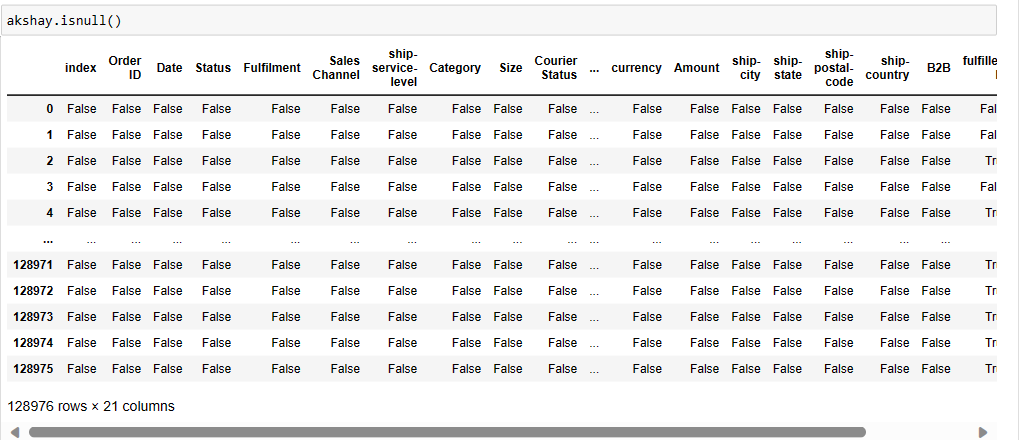
**STEP 6:** CHECKING THE DATA-TYPES OF THE DATA FRAME



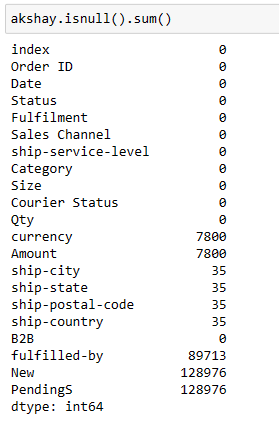
**STEP 7**: CHECKING THE COLUMNS NAME OF THE DATA-FRAME



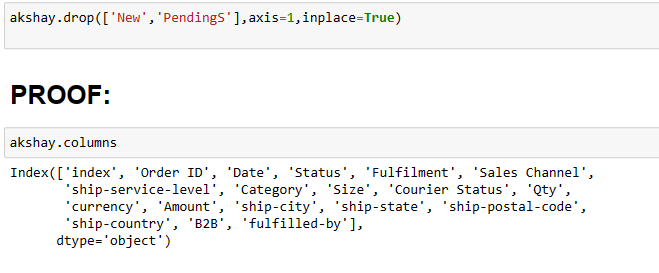
**STEP 8:** CHECKING THE NULLVALUES IN THE DATA FRAME



**STEP 9:** FETCHING THE SUM OF NULL VALUES



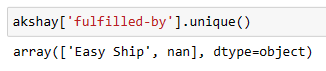
**STEP 10**: REMOVING THE COLUMNS NAME OF THE DATA-FRAME , BASED ON OUR REQUIREMENT.

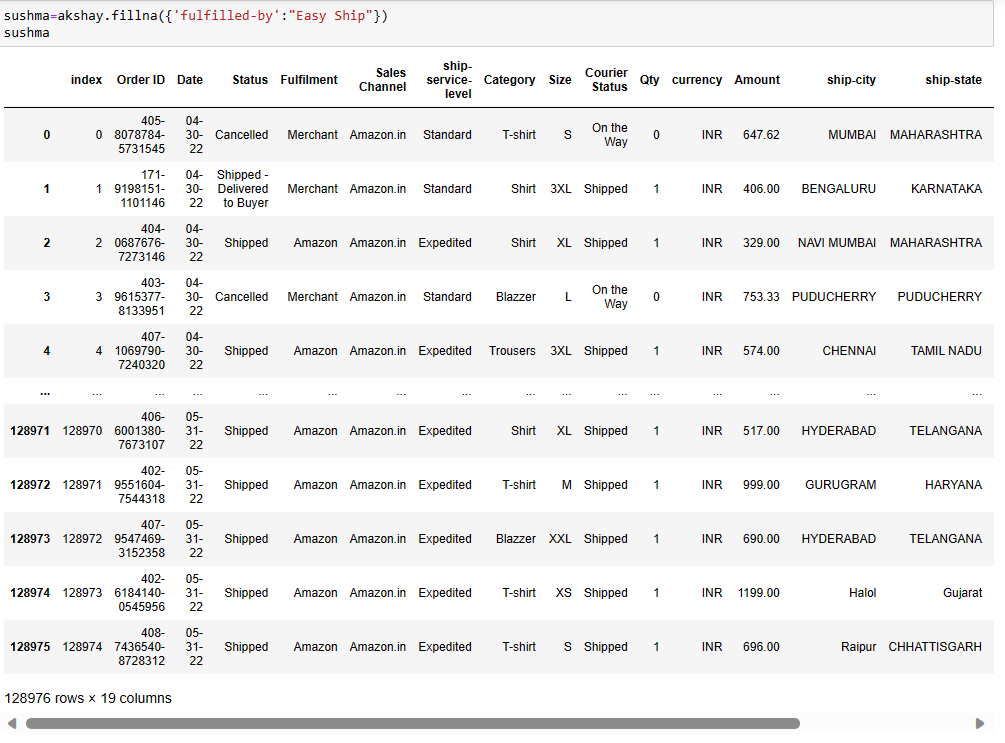


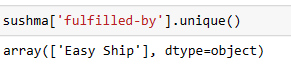
**STEP 11:** FILLNA PROCESS

FIRST WE ARE FETCHING UNIQUE VALUES FOR FILLNA PROCESS , MOST COMMONLY REPEATED VALUES .

FOR NAN VALUES WE ARE FILLING WITH “EASY SHIP” VALUE USING FILLNA FUNCTION ,FOR THE COLUMN “FULFILLED-BY”.

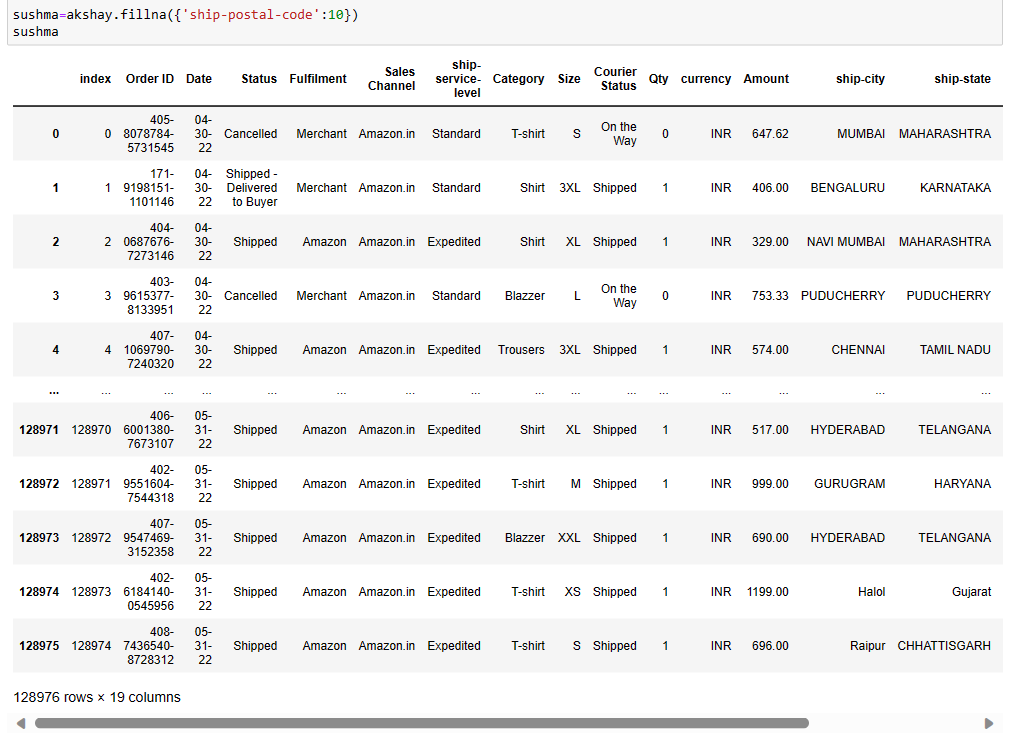






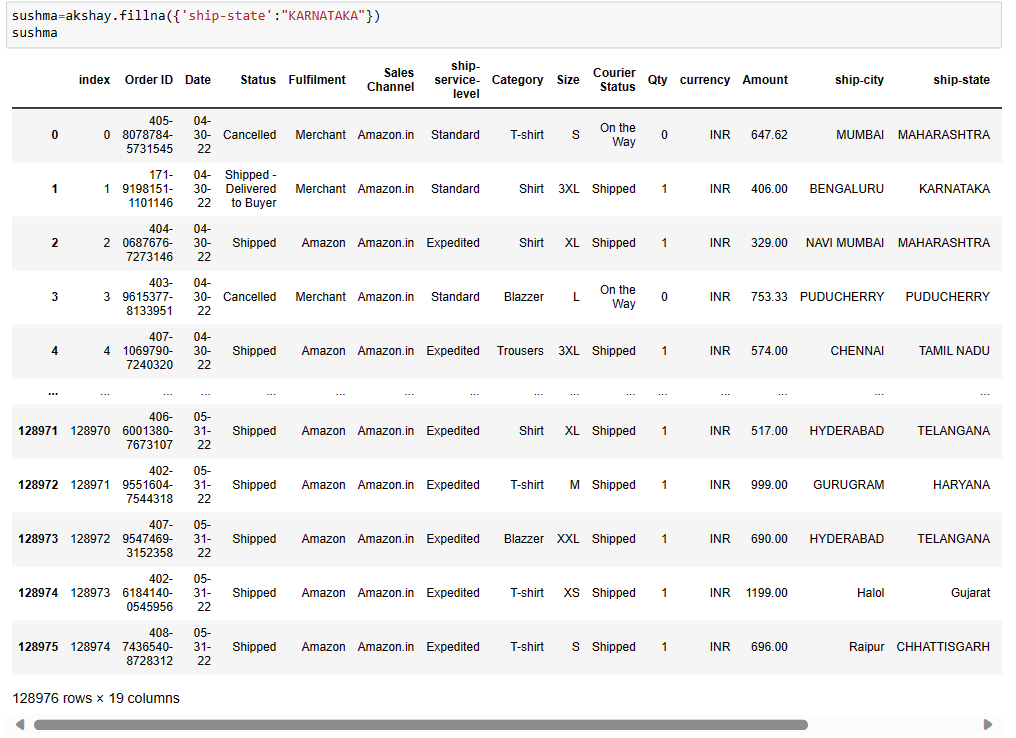
FIRST WE ARE FETCHING UNIQUE VALUES FOR FILLNA PROCESS , MOST COMMONLY REPEATED VALUES .

FOR NAN VALUES WE ARE FILLING WITH “10” VALUE USING FILLNA FUNCTION ,FOR THE COLUMN “SHIP-POSTAL-CODE”.



FIRST WE ARE FETCHING UNIQUE VALUES FOR FILLNA PROCESS , MOST COMMONLY REPEATED VALUES .

FOR NAN VALUES WE ARE FILLING WITH “KARNATAKA” VALUE USING FILLNA FUNCTION ,FOR THE COLUMN “SHIP-STATE”.

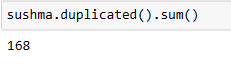


**STEP 12:** AFTER PROCESS , FETCHING SUM OF NULL VALUES.

AFTER THE REMOVAL OF NAN VALUES OR REPLACE NAN VALUES WITH UR SPECIFIED VALUES ..WE ARE FETCHING SUM OF NULL VALUES FOR DIFFERENT COLUMNS .

**STEP 13:** DUPLICATED VALUES

FETCHING DUPLICATED VALUES SUM FOR REMOVING DUPLICATED VALUES .



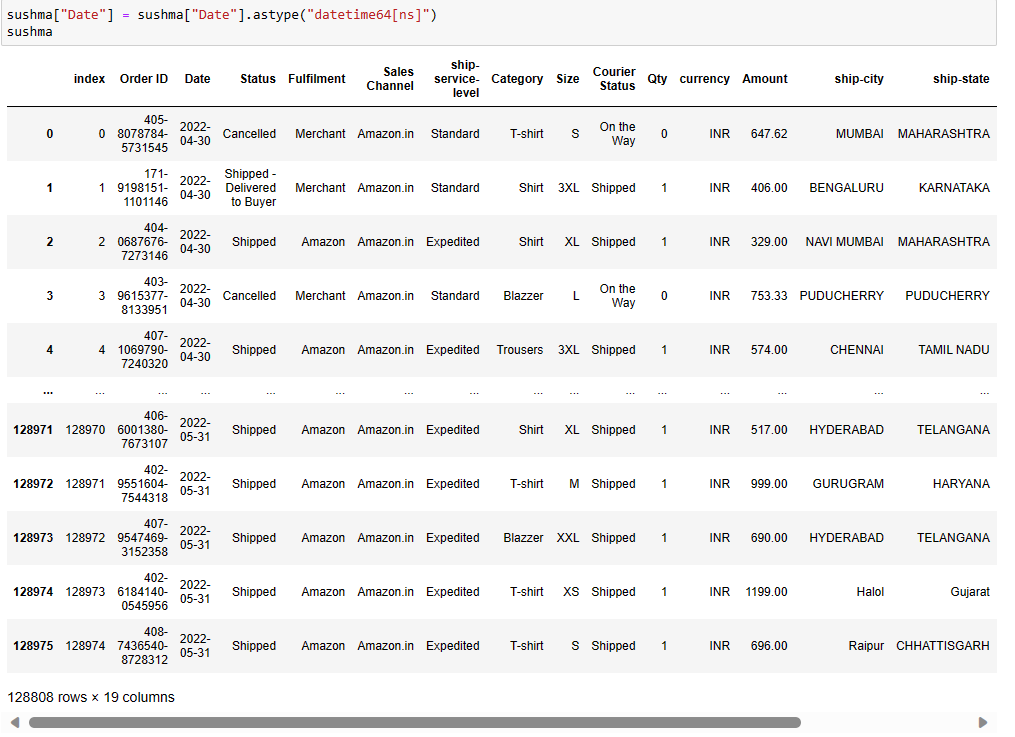
**STEP 14:** CLEARING THE DUPLICATED VALUES

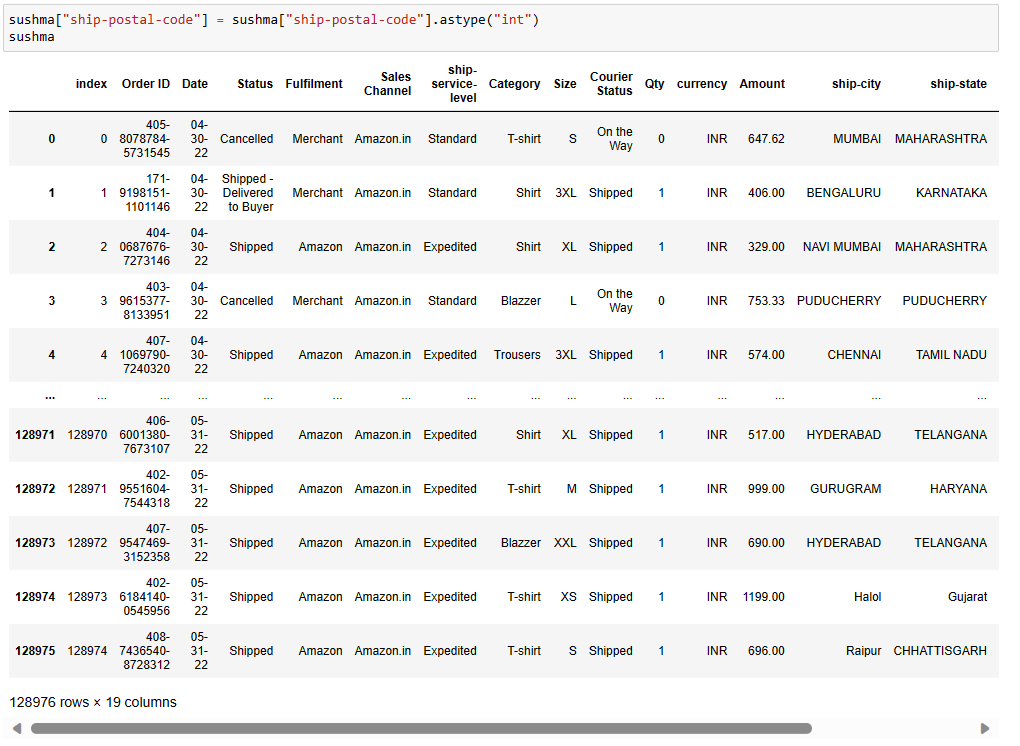
AFTER FETCHING DUPLICATED VALUES SUM, WE WILL DROP THE DUPLICATED VALUES USING DROPING METHOD …

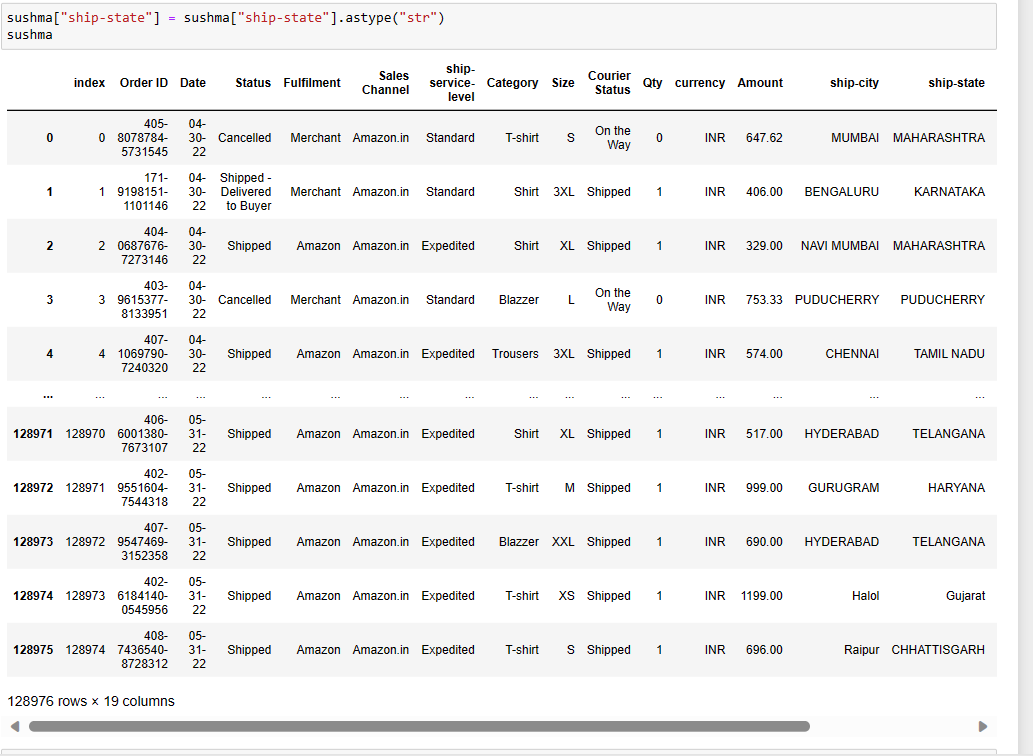


**STEP 15:** CONVERTING DATA-TYPES

DATA-TYPE CONVERSION IS THE PROCESS OF CONVERTING A DATA TYPE INTO ANOTHER DATA TYPE. IMPLICIT TYPE CONVERSION IS PERFORMED BY A PYTHON INTERPRETER ONLY.







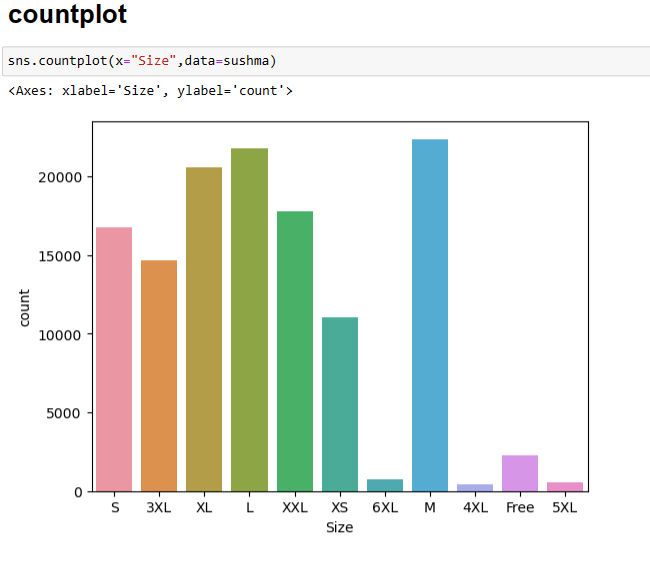
**STEP 16**: GRAPHICAL REPRESENTATION .

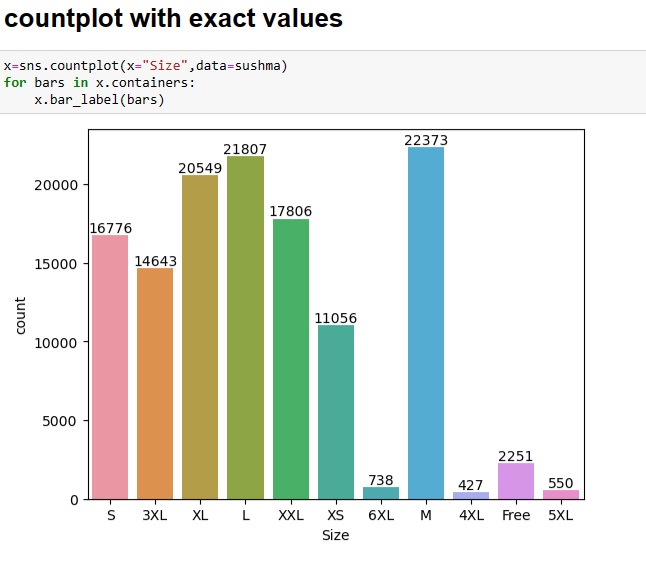
**DATA VISUALIZATION**

DATA VISUALIZATION IS THE REPRESENTATION OF DATA THROUGH USE OF COMMON GRAPHICS, SUCH AS CHARTS, PLOTS, INFOGRAPHICS, AND EVEN ANIMATIONS. THESE VISUAL DISPLAYS OF INFORMATION COMMUNICATE COMPLEX DATA RELATIONSHIPS AND DATA-DRIVEN INSIGHTS IN A WAY THAT IS EASY TO UNDERSTAND.

**COUNT-PLOT:**

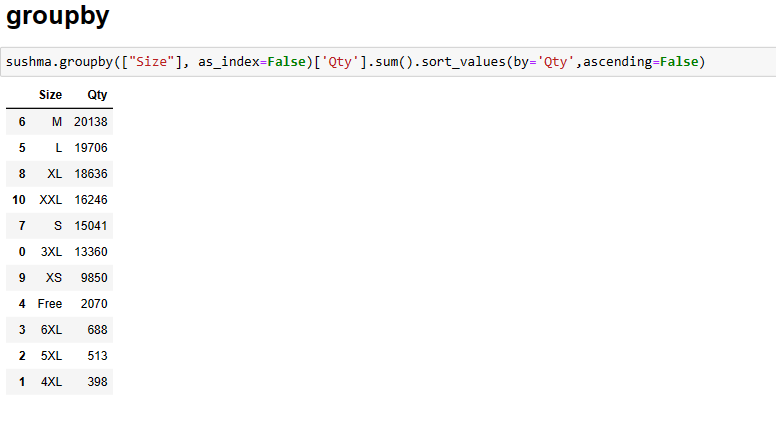
Count Plot in Seaborn is used to display the counts of observations in each categorical bin using bars.

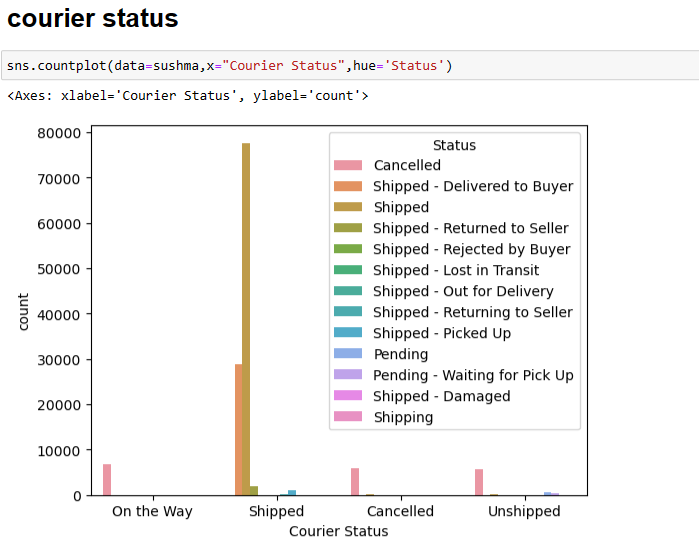


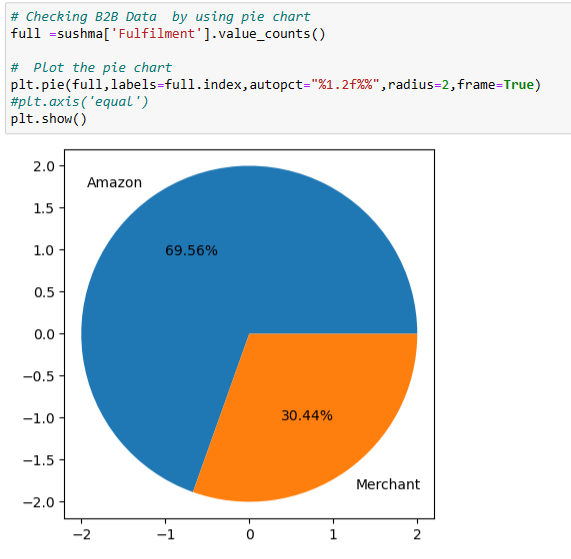


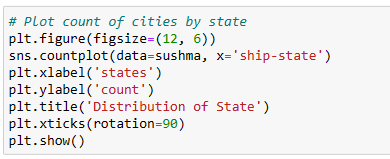
**GROUP-BY:**

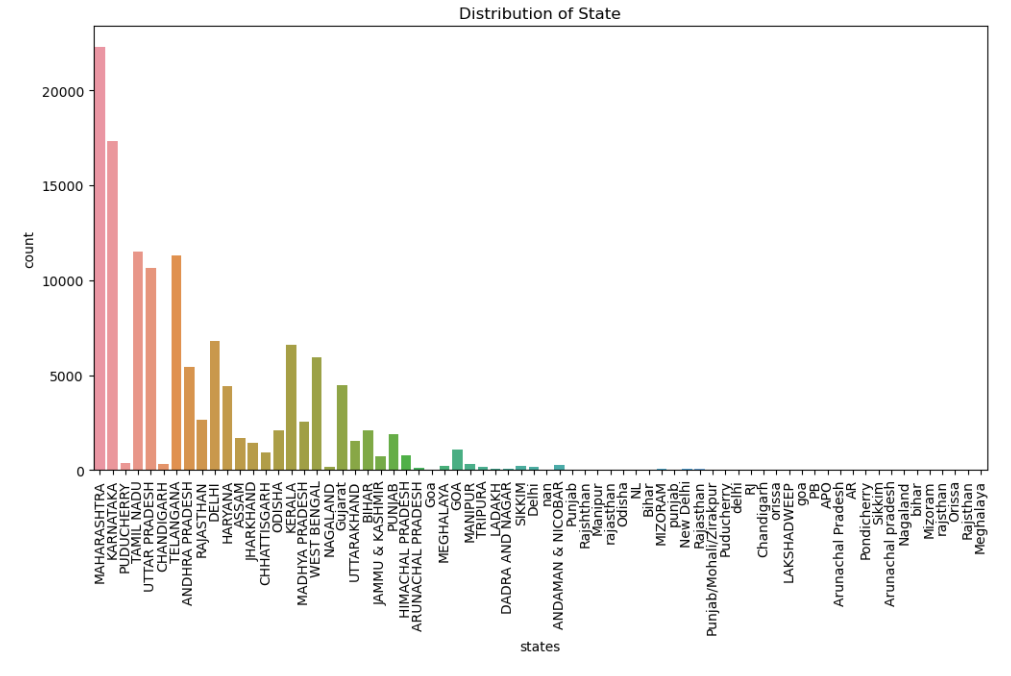
A GROUPBY OPERATION INVOLVES SOME COMBINATION OF SPLITTING THE OBJECT, APPLYING A FUNCTION, AND COMBINING THE RESULTS. THIS CAN BE USED TO GROUP LARGE AMOUNTS OF DATA AND COMPUTE OPERATIONS ON THESE GROUPS. PARAMETERS BYMAPPING, FUNCTION, LABEL, PD.GROUPER OR LIST OF SUCH. USED TO DETERMINE THE GROUPS FOR THE GROUPBY.

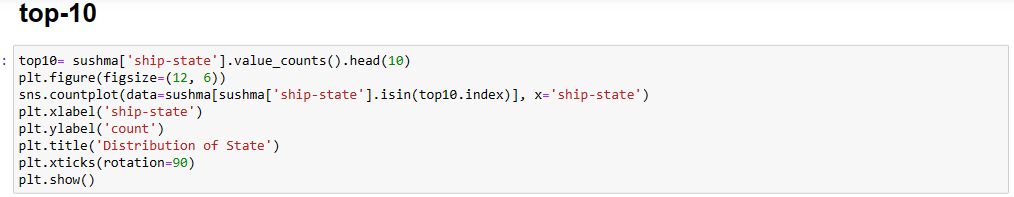


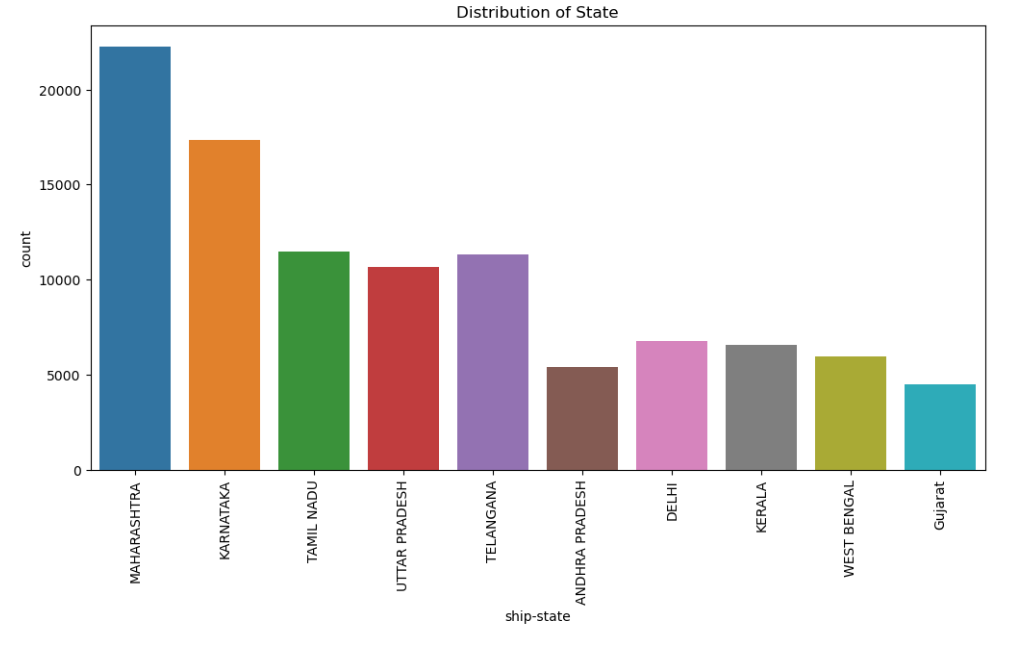












S

conclusion

the data analysis reveals that the business has a significant customer base in Maharashtra state, mainly serves retailers,fulfills orders through Amazon ,experiences high demand for T-shirts and sees M-size as the preferred choice among buyers.[¶](http://localhost:8888/notebooks/AKSHAY%20AMAZON%20PROJECT%20DOCUMENTATION%20%20.ipynb#the-data-analysis-reveals-that-the-business-has-a-significant-customer-base-in-Maharashtra-state,-mainly-serves-retailers,fulfills-orders-through-Amazon-,experiences-high-demand-for-T-shirts-and-sees-M-size-as-the-preferred-choice-among-buyers.)

**UNDER THE GUIDANCE OF :**

**SAI CHARAN**

TECHINCAL HEAD

EINFACHZAUBERER TECHNOLOGIES

PVT LTD, BANGALORE.