**PROJECT TITLE: PRODUCT SALES ANALYSIS**

**DAC\_PHASE 3 Submission document**

Team Leader : S.Sharmila / au812121106030

Sales analytics is the practice of generating insights from sales data, trends, and metrics to set targets and forecast future sales performance. Sales analysis is mining your data to evaluate the performance of your sales team against its goals. It provides insights about the top performing and underperforming products/services, the problems in selling and market opportunities, sales forecasting, and sales activities that generate revenue.

**Content**

* Order ID - An Order ID is the number system that Amazon uses exclusively to keep track of orders. Each order receives its own Order ID that will not be duplicated. This number can be useful to the seller when attempting to find out certain details about an order such as shipment date or status.
* Product - The product that have been sold.
* Quantity Ordered - Ordered Quantity is the total item quantity ordered in the initial order (without any changes).
* Price Each - The price of each products.
* Order Date - This is the date the customer is requesting the order be shipped.
* Purchase Address - The purchase order is prepared by the buyer, often through a purchasing department. The purchase order, or PO, usually includes a PO number, which is useful in matching shipments with purchases; a shipping date; billing address; shipping address; and the request items, quantities and price.

**Target**

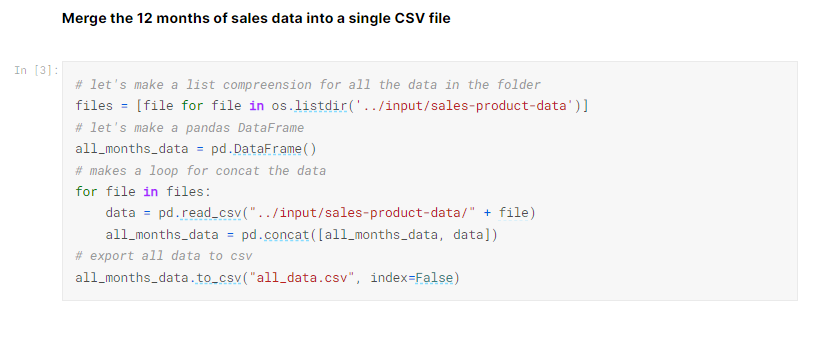
A target market analysis is an assessment of how your product or service fits into a specific market and where it will gain the most.

**Task:**

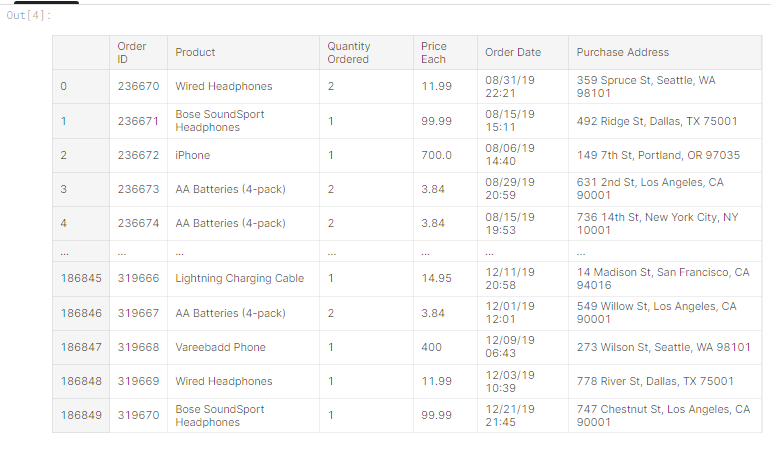
* Q: What was the best Year for sales? How much was earned that Year?
* Q: What was the best month for sales? How much was earned that month?
* Q: What City had the highest number of sales?
* Q: What time should we display adverstisement to maximize likelihood of customer's buying product?
* Q: What products are most often sold together?
* Q: What product sold the most? Why do you think it sold the most?

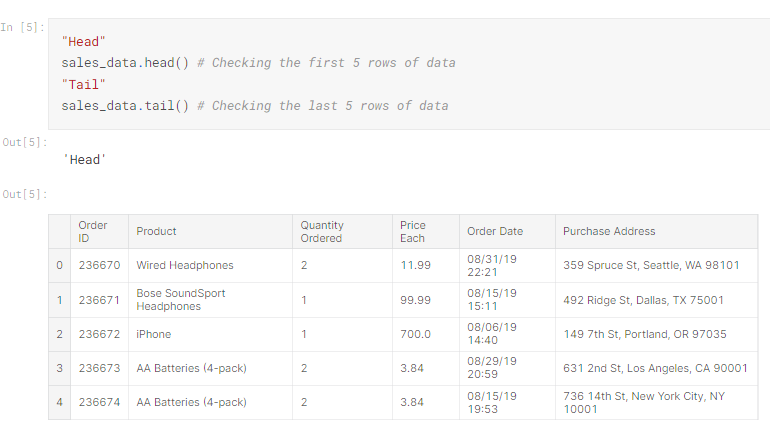
**How Much Probability?**

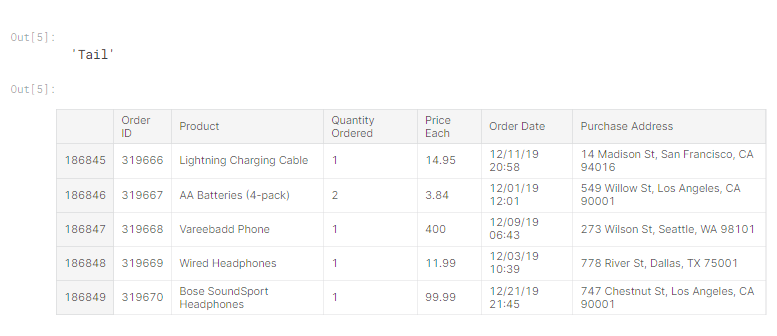
* How much probability for next people will ordered USB-C Charging Cable?
* How much probability for next people will ordered iPhone?
* How much probability for next people will ordered Google Phone?
* How much probability other peoples will ordered Wired Headphon





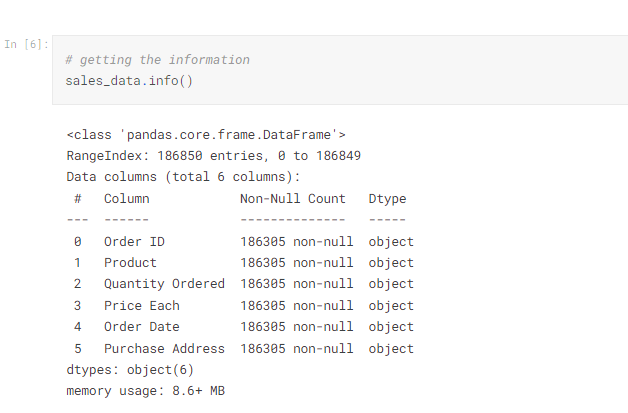


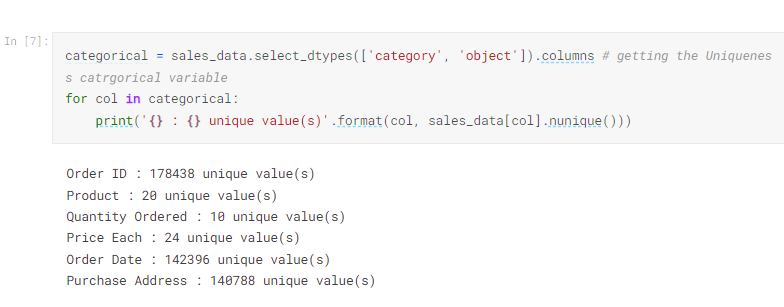


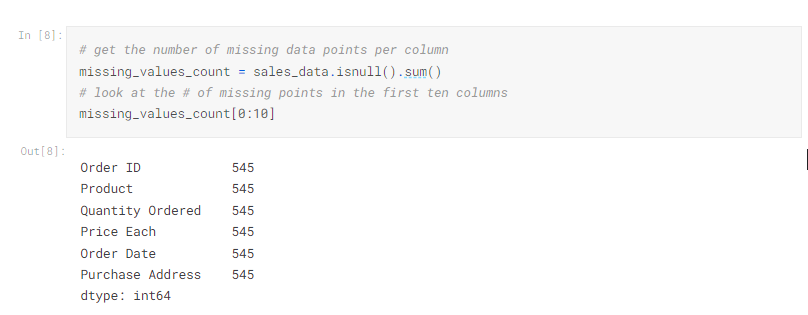


# ****Data Preprocessing****

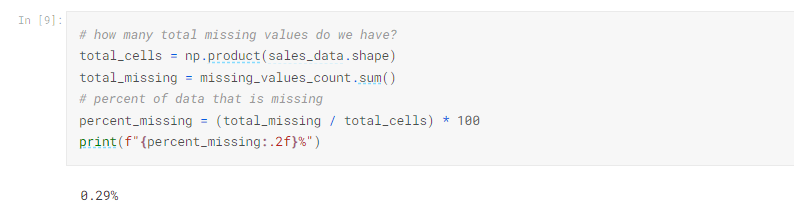
Data preprocessing can refer to manipulation or dropping of data before it is used in order to ensure or enhance performance, and is an important step in the data mining process. The phrase "garbage in, garbage out" is particularly applicable to data mining and machine learning projects.







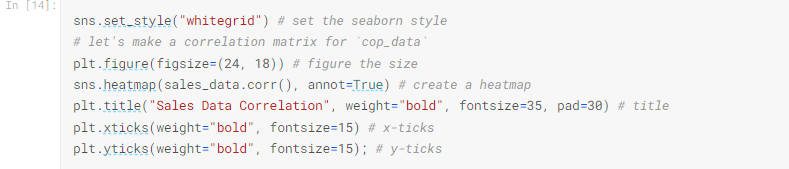
It might be helpful to see what percentage of the values in our dataset were missing to give us a better sense of the scale of this problem:

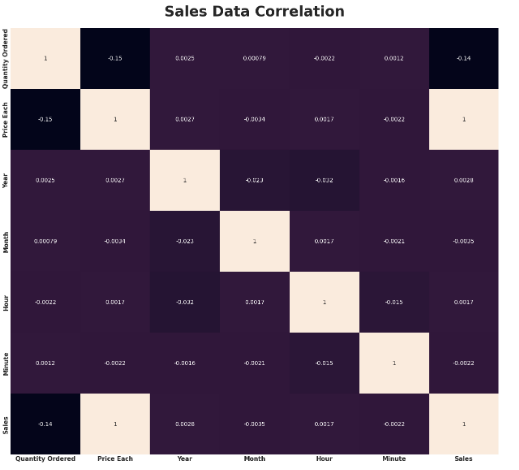


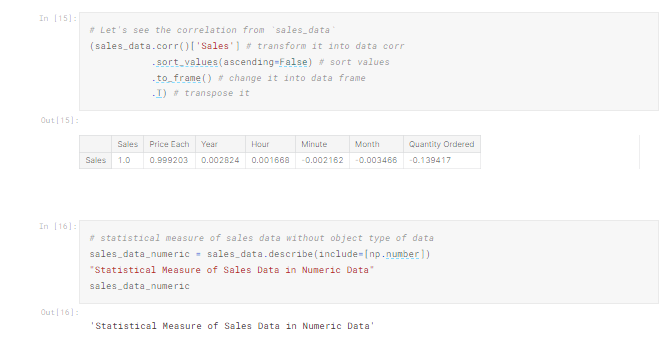
Data Analysis

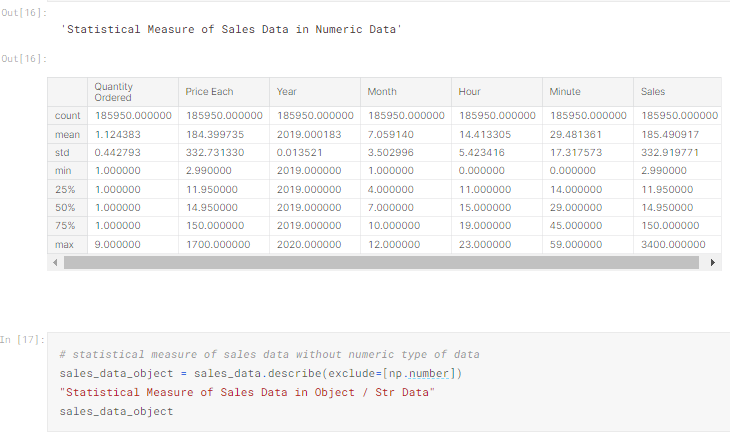
Data Analysis is the process of systematically applying statistical and/or logical techniques to describe and illustrate, condense and recap, and evaluate data. Indeed, researchers generally analyze for patterns in observations through the entire data collection phase (Savenye, Robinson, 20042004). analyze and investigate data sets and summarize their main characteristics, often employing data visualization methods.

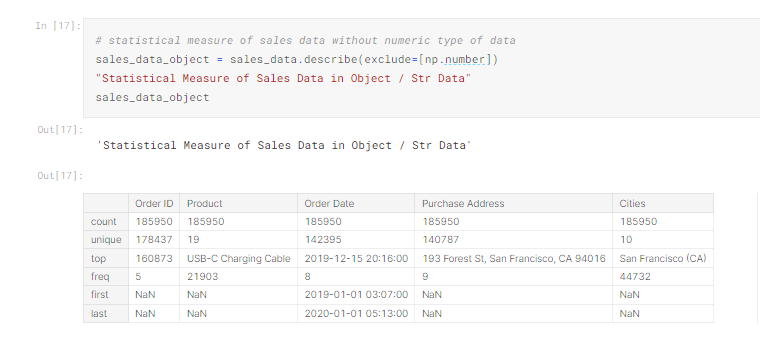
Or, the easier, you can say in Data Analysis we (Data Scientist or Data Analyst) what ever you want to call that, in this section, we're looking for the correlation and also the relationships between every data (features and labels) or the variables using and applying the statistical and visualization methods for looking some patterns.











## ****Quantity Ordered****

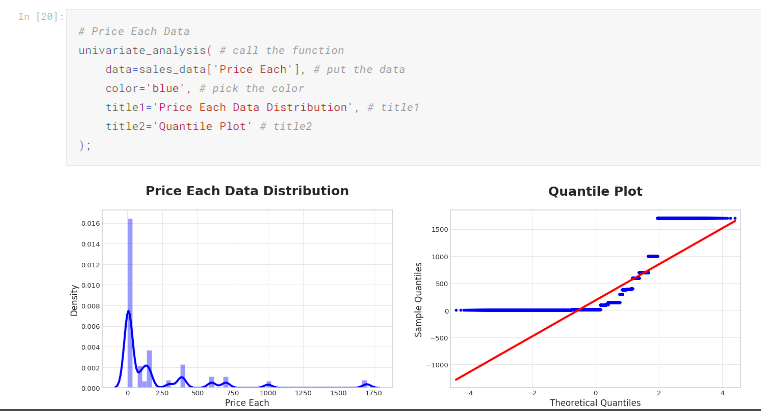
Find the proportion that lies in between two standard deviation (σ�) from mean (μ�), and let's try to interprete that. and In the Quantity Ordered Data, the μ=1.12�=1.12 and the σ=0.44�=0.44, then without further ado let's calculate it.

#### ****Calculation:****

* 1.12−2(0.44)=0.21.12−2(0.44)=0.2
* 1.12+2(0.44)=21.12+2(0.44)=2

#### Interpretation:

At least 75%75% of the Sales Data Quantity Ordered population in the USA has a Quantity Ordered range from 0−20−2 item/product.



## ****Price Each****

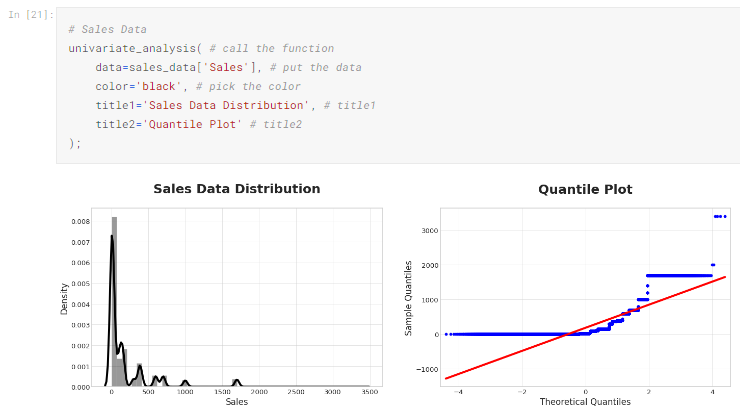
Find the proportion that lies in between two standard deviation (σ�) from mean (μ�), and let's try to interprete that. and in the Price Each Data, the μ=184.3�=184.3 and the σ=332.7�=332.7, then without further ado let's calculate it.

#### ****Calculation:****

* 184.3−2(332.7)=−481184.3−2(332.7)=−481
* 184.3+2(332.7)=849.7184.3+2(332.7)=849.7

#### Interpretation:

At least 75%75% of the population Sales Price data for each item/product in the USA has a price range for each item/product from 0−849.70−849.7 (USD).



## ****Sales****

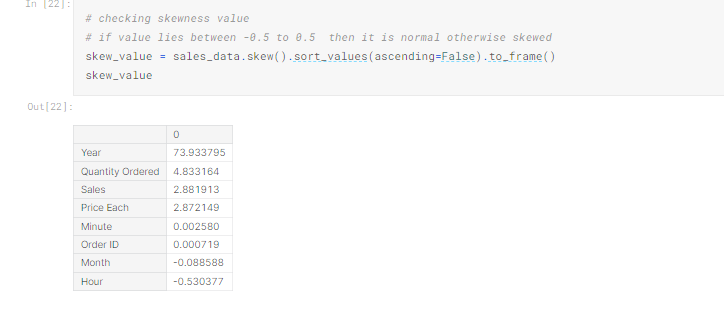
Find the proportion that lies in between two standard deviation (σ�) from mean (μ�), and let's try to interprete that. and in the Sales Data, the μ=185.4�=185.4 and the σ=332.9�=332.9, then without further ado let's calculate it.

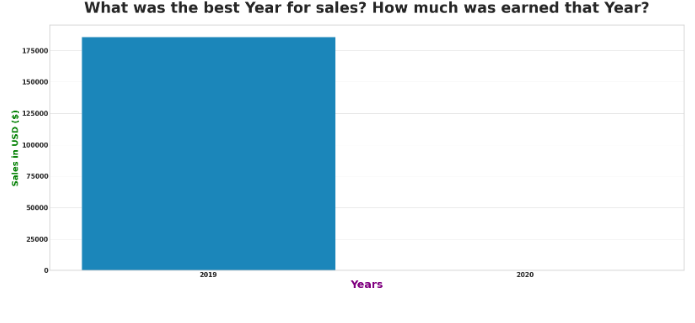
#### ****Calculation:****

* 185.4−2(332.9)=−480185.4−2(332.9)=−480
* 185.4+2(332.9)=851.19185.4+2(332.9)=851.19

#### Interpretation:

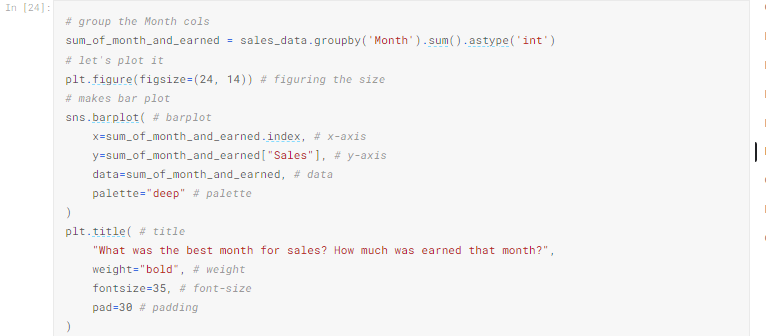
At least 75%75% of population Sales Data customers in USA have Sales range from 0−851.190−851.19 (USD)



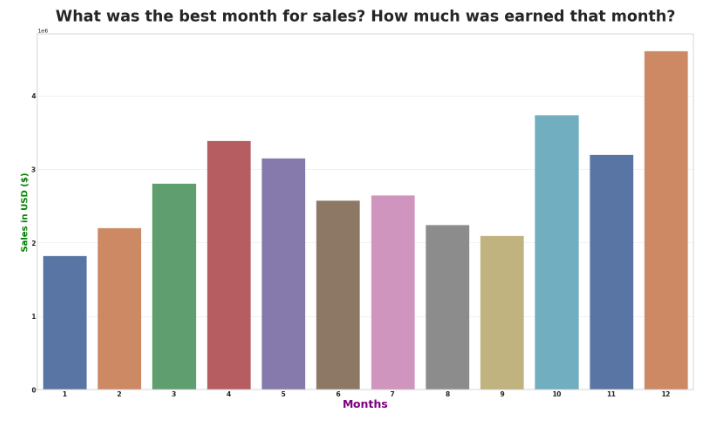


### Answer:

When viewed from the data above, 20192019 was the best year that had the highest number of sales, which was $34,483,365$34,483,365, compared to 20202020 which only had $8,670$8,670 in sales, this is due to the lack of data in 20202020 which caused a data imbalance.

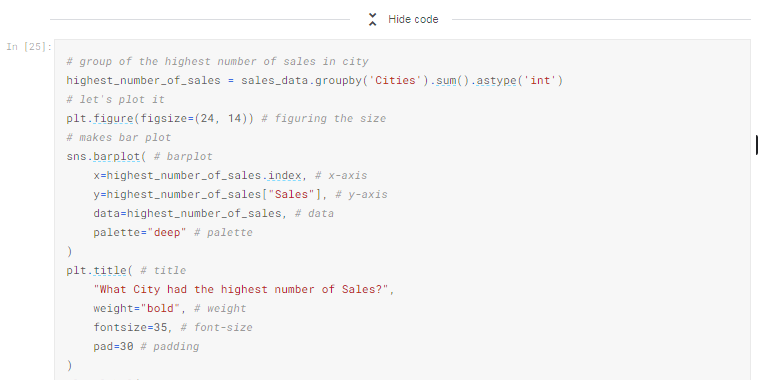


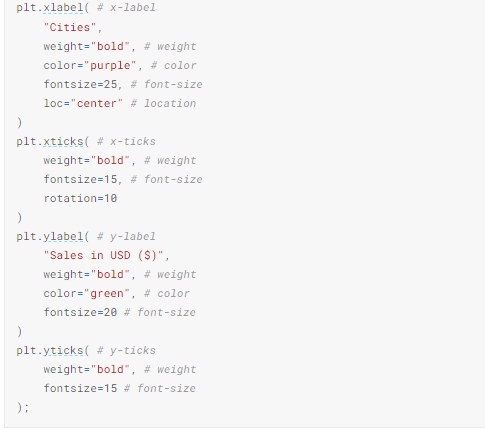


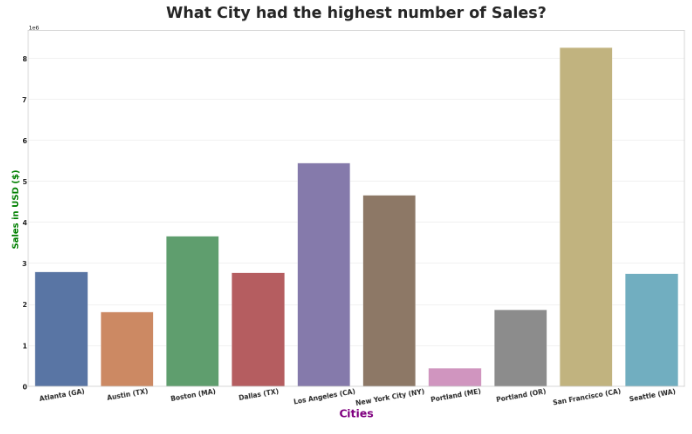


### Answer:

The best month to sell is shown in the visualization above is December which has a record number of sales reaching $4,613,443$4,613,443, sales,This may be because in December there is Christmas, where many people buy groceries to make cakes or toys as gifts for loved ones.

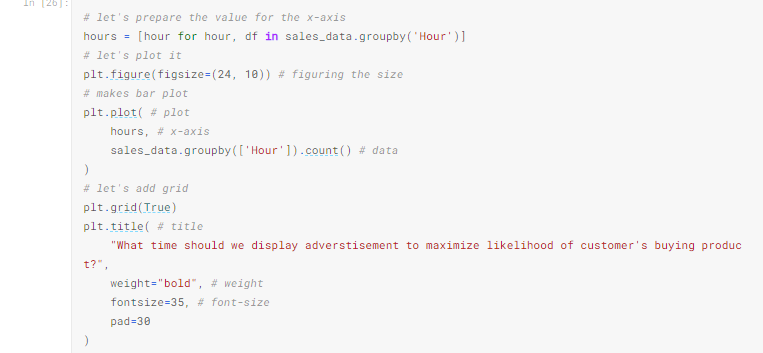






### Answer:

The city that has the most sales data in the above visualization is San Francisco, with total sales reaching $8,262,203$8,262,203





### ****How much probability for next people will order Wired Headphones?****

