```
/*============== PROJECT - SALES ANALYSIS ===============*/

/*=================================================================*/
/* Creating path macro variable for project-data folder */
/*=================================================================*/

%let path=/home/u58427518/sas_submissions/data/project/;

/*=================================================================*/
/* Importing orders.csv file */
/*=================================================================*/

proc import datafile= "&path/orders.csv"
out = orders
dbms = csv
replace
;
run;

/*=================================================================*/
/* Importing country_lookup file */
/*=================================================================*/

libname cntry xlsx "&path/country_lookup.xlsx";


/*=================================================================*/
/* Viewing first ten records of orders.csv */
/*=================================================================*/

proc print data=orders(obs=10);
run;
```

Obs	Order_ID	Order_Date	Delivery_Date	Order_Type	Product_ID	Product_Line	Product_Category	Quantity	Retail_Price	Cost_Price	Customer_Country	Customer_Continent	Customer_Dob
1	1244336610	31DEC2018	04JAN2019	3	230101000000	Outdoors	Outdoors	3	544.5	228.9	RU	Asia	1980-06-02
2	1244336421	31DEC2018	04JAN2019	1	230100000000	Outdoors	Outdoors	2	283.4	122	NL	Europe	2000-09-08
3	1244336321	31DEC2018	04JAN2019	2	220200000000	Clothes & Shoes	Shoes	2	241.4	133	NL	Europe	2002-07-31
4	1244336156	31DEC2018	31DEC2018	1	240500000000	Sports	Running - Jogging	-1	117.8	52.2	IT	Europe	1995-04-23
5	1244336082	31DEC2018	31DEC2018	1	220200000000	Clothes & Shoes	Shoes	2	80	40.2	IT	Europe	1992-04-09
6	1244335617	31DEC2018	31DEC2018	1	210201000000	Children	Children Sports	2	37	16.5	IT	Europe	1974-12-22
7	1244335593	31DEC2018	31DEC2018	1	220200000000	Clothes & Shoes	Shoes	2	315.6	158	IT	Europe	1987-08-16
8	1244335565	31DEC2018	31DEC2018	99	220101000000	Clothes & Shoes	Clothes	2	23	7.5	IT	Europe	1975-08-04
9	1244335244	31DEC2018	31DEC2018	99	240200000000	Sports	Golf	2	122	59.9	GB	Europe	1980-07-09
10	1244334892	31DEC2018	31DEC2018	1	220200000000	Clothes & Shoes	Shoes	4	450.4	205.2	FR	Europe	1960-01-14

```
/*============================================================*/
/* Viewing first ten records of country_lookup.xlsx */
/*============================================================*/

proc print data=cntry.countries(obs=10);
run;
/*============================================================*/
```

Obs	Country_Key	Lat	Lon	Country_Name
1	AD	42.546245	1.601554	Andorra
2	AE	23.424076	53.847818	United Arab Emirates
3	AF	33.93911	67.709953	Afghanistan
4	AG	17.060816	-61.796428	Antigua/Barbuda
5	AG	17.060816	-61.796428	Antigua/Barbuda
6	AI	18.220554	-63.068615	Anguilla
7	AL	41.153332	20.168331	Albania
8	AM	40.069099	45.038189	Armenia
9	AN	12.226079	-69.060087	Netherlands Antilles
10	AO	-11.202692	17.873887	Angola

```
/*============================================================*/
/* Checking length and data-type of all variables in country data
/*============================================================*/

country_lookup.xlsx */
proc contents data=cntry.countries;
run;
```

Alphabetic List of Variables and Attributes						
#	Variable	Type	Len	Format	Informat	Label
1	Country_Key	Char	2	$2.	$2.	Country_Key
4	Country_Name	Char	24	$24.	$24.	Country_Name
2	Lat	Num	8	BEST.		Lat
3	Lon	Num	8	BEST.		Lon

```
/*============================================================*/
/* All the variables have  proper data-type and length */
/*============================================================*/
```

```
/*============================================================*/
/* Checking length and data-type of all variables in orders.csv */
/*============================================================*/

proc contents data=orders;
run;
```

Alphabetic List of Variables and Attributes					
#	Variable	Type	Len	Format	Informat
10	Cost_Price	Num	8	BEST12.	BEST32.
12	Customer_Continent	Char	6	$6.	$6.
11	Customer_Country	Char	2	$2.	$2.
13	Customer_Dob	Num	8	YYMMDD10.	YYMMDD10.
3	Delivery_Date	Num	8	DATE9.	DATE9.
2	Order_Date	Num	8	DATE9.	DATE9.
1	Order_ID	Num	8	BEST12.	BEST32.
4	Order_Type	Num	8	BEST12.	BEST32.
7	Product_Category	Char	24	$24.	$24.
5	Product_ID	Num	8	BEST12.	BEST32.
6	Product_Line	Char	15	$15.	$15.
8	Quantity	Num	8	BEST12.	BEST32.
9	Retail_Price	Num	8	BEST12.	BEST32.

```
/*============================================================*/
/* All the variables have proper data-type and length */
/*============================================================*/
```

```
/*=============================================================*/
          /* Exploring and validating the data */
/*=============================================================*/

/*=============================================================*/
/* Validating Categorical Data (country_lookup data):
/*=============================================================*/

--- Validation(1):

--- Each country key should have only one country name assigned to
it.
--- After using 'proc freq', it was found that some country keys
have been assigned multiple
country names.
--- Upon further analysis it was found that the issue resided in
the country names being slightly different for some countries
--- For example:
--- For country_key = CF, country_name = Central African Rep. as
well as country_name = Central African Republic
--- For country_key = GB, country_name = United Kingdom as well as
country_name = Great Britain

--- Fix:

--- Keeping the first occurence for a country_key and its
     respective country_name
--- Removing second/duplicate country_name occurence for same
     country_key.
--- Saving the cleaned data-set with duplicate entries removed in
     another location.
--- Saving the duplicates in another output dataset
--- Making sure that only invalid records were removed.

*/
```

```
/*=============================================================*/
     /* Analyzing categorical variable in countries-sheet */
/*=============================================================*/

proc freq data=cntry.countries order=freq;
tables country_key country_name;
```

Country_Name				
Country_Name	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Antigua/Barbuda	2	0.84	2	0.84
Afghanistan	1	0.42	3	1.27
Albania	1	0.42	4	1.69
Algeria	1	0.42	5	2.11
American Samoa	1	0.42	6	2.53
Andorra	1	0.42	7	2.95
Angola	1	0.42	8	3.38
Anguilla	1	0.42	9	3.80
Antarctica	1	0.42	10	4.22
Argentina	1	0.42	11	4.64
Armenia	1	0.42	12	5.06
Aruba	1	0.42	13	5.49
Australia	1	0.42	14	5.91
Austria	1	0.42	15	6.33
Azerbaijan	1	0.42	16	6.75
Bahamas	1	0.42	17	7.17
Bahrain	1	0.42	18	7.59
Bangladesh	1	0.42	19	8.02
Barbados	1	0.42	20	8.44

Country_Key				
Country_Key	Frequency	Percent	Cumulative Frequency	Cumulative Percent
AG	2	0.84	2	0.84
CF	2	0.84	4	1.69
GB	2	0.84	6	2.53
US	2	0.84	8	3.38
AD	1	0.42	9	3.80
AE	1	0.42	10	4.22
AF	1	0.42	11	4.64
AI	1	0.42	12	5.06
AL	1	0.42	13	5.49
AM	1	0.42	14	5.91
AN	1	0.42	15	6.33
AO	1	0.42	16	6.75
AQ	1	0.42	17	7.17
AR	1	0.42	18	7.59
AS	1	0.42	19	8.02
AT	1	0.42	20	8.44
AU	1	0.42	21	8.86
AW	1	0.42	22	9.28
AZ	1	0.42	23	9.70
BA	1	0.42	24	10.13
BB	1	0.42	25	10.55

```
/*=============================================================*/
/* Further analyzing data for countries with duplicate keys */
/*=============================================================*/

proc print data=cntry.countries;
where country_key in ('AG','CF','GB','US');
var country_key country_name;
```

Obs	Country_Key	Country_Name
1	AG	Antigua/Barbuda
2	AG	Antigua/Barbuda
3	CF	Central African Rep.
4	CF	Central African Republic
5	GB	United Kingdom
6	GB	Great Britain
7	US	United States
8	US	United States of America

** Found aboverecords with duplicate country keys **

```
/*=============================================================*/
/* Removing duplicate country_name assigned for some country_key*/
/*=============================================================*/

proc sort data=cntry.countries out=country_clean
dupout = dups
nodupkey;
by country_key;
/*=============================================================*/
/* Checking the duplicate records that were removed */
/*=============================================================*/

proc print data=dups;
run;
```

Obs	Country_Key	Lat	Lon	Country_Name
1	AG	17.060816	-61.796428	Antigua/Barbuda
2	CF	6.611111	20.939444	Central African Republic
3	GB	55.378051	-3.435973	Great Britain
4	US	37.09024	-95.712891	United States of America

```
/* 4 records with duplicate keys removed and stored separately */
```

```
/*================================================================*/
/* Categorical Data Analysis (Orders Data) :
/*================================================================*/

--- Validation (1):
--- For some of the records it was found that the order date was
     later than the delivery date.
--- That is invalid, since first the order must be placed only
     then the delivery can take place.

--- Fix (1):
--- Filtering only those records where delivery date is later
--- or same as order date.

--- Validation (2):
--- Item Types : 1-Retail, 2-Phone, 3-Internet, 99-Invalid.
--- Some of the items have item-type as "99", which is invalid.

--- Fix (2):
--- Creating new column "Order Type Detail"
--- With following classes: 1-Retail, 2-Phone, 3-Internet, 4-
     Unknown

--- Validation (3):
--- Validating customer_country and continents columns
--- Some  country keys are in lowercase
--- Such country keys won't match with country keys in
     country_lookup data.

--- Validation (4):
--- Validating data for customer_continent
--- Checking if all records have no continents beside 5 valid ones
*/
```

```
/*=================================================================*/
/* Creating PDF file to print all anomalies found */
/*=================================================================*/
ods pdf file="&path.\anomalies.pdf";

/*=================================================================*/
/* Validatingfor records that have order_date later than delivery
date */
/*=================================================================*/

title 'Orders With Invalid Date';
proc print data=orders;
where order_date > delivery_date;
run;
```

Orders With Invalid Date

Obs	Order_ID	Order_Date	Delivery_Date	Order_Type	Product_ID	Product_Line	Product_Category	Quantity	Retail_Price	Cost_Price	Customer_Country	Customer_Continent	Customer_Dob
1058	1244098378	04DEC2018	01JAN1960	2	220200000000	Clothes & Shoes	Shoes	2	239.6	104.1	US	North	1983-02-01
2467	1243695400	21OCT2018	01JAN1960	1	220100000000	Clothes & Shoes	Clothes	1	57.3	28.75	NL	Europe	1984-02-25
3292	1243440852	23SEP2018	01JAN1960	1	240100000000	Sports	Assorted Sports Articles	1	419	209.45	FR	Europe	1985-11-10
4462	1243061668	12AUG2018	01JAN1960	1	230100000000	Outdoors	Outdoors	1	176	74.55	NL	Europe	1991-02-28
5975	1242634685	26JUN2018	01JAN1960	1	220100000000	Clothes & Shoes	Clothes	1	25.6	12.1	FR	Europe	1967-05-31

```
/* Found : 5 records with anomaly*/

/*=================================================================*/
/* Validating for records that have item-type as "99" (Invalid) */
/*=================================================================*/

title 'Orders With Invalid Order Type';
proc freq data=orders;
table order_type;
run;
```

Orders With Invalid Order Type

The FREQ Procedure

Order_Type	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	8337	77.29	8337	77.29
2	1090	10.11	9427	87.40
3	1356	12.57	10783	99.97
99	3	0.03	10786	100.00

```
/* Found : 3 records with anomaly */
```

```
/*============================================================*/
/* Validating country-keys for the orders data */
/*============================================================*/

title 'Orders With Invalid Case Of The Country Key';
proc freq data=orders;
table customer_country;
run;
```

Customer_Country	Frequency	Percent	Cumulative Frequency	Cumulative Percent
AE	2	0.02	2	0.02
AT	12	0.11	14	0.13
AU	731	6.78	745	6.91
au	1	0.01	10783	99.97
be	2	0.02	10785	99.99
fr	1	0.01	10786	100.00

```
/* Found : 3 keys with anomaly */

/*============================================================*/
/* Closing PDF */
/*============================================================*/

ods pdf close;

* OUTPUT FOR ANOMALIES HAS BEEN STORED IN anomalies.pdf *

/*============================================================*/
/* Validating customer_continent for the orders data */
/*============================================================*/

proc freq data=orders;
table customer_continent;
```

Customer_Continent	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Africa	14	0.13	14	0.13
Asia	12	0.11	26	0.24
Austra	732	6.79	758	7.03
Europe	7598	70.44	8356	77.47
North	2430	22.53	10786	100.00

```
/* Found : No anomalies */
```

```
/*=============================================================*/
          * Analyzing Numeric Data  (Orders Data) *
/*=============================================================*/

--- Validation (1):
--- Analyzing Quantity, Retail_Price and Cost_Price columns
--- Using 'proc means' --- finding -- mean, median, std. dev.

--- Column Quantity has minimum value = (-1)
--- That is invalid as minimum value should be 0.

--- Fix (1):
--- Finding records for which quantity has min. value = -1
--- Replacing quantity value for such records with '0'.

/*=============================================================*/
/* Performing statistical analysis of quantitative columns */
/*=============================================================*/

proc means data=orders;
var quantity retail_price cost_price;
```

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
Quantity	10786	1.6802336	0.8965046	-1.0000000	8.0000000
Retail_Price	10786	139.6040413	183.5431780	0.6300000	3191.00
Cost_Price	10786	65.4603662	89.0741604	0.2000000	1583.60

/* Found : 1 Anomaly */

```
/*=============================================================*/
/* Finding records where quantity is less than '0' or negative */
/*=============================================================*/

proc print data=orders;
where quantity < 0;
```

Obs	Order_ID	Order_Date	Delivery_Date	Order_Type	Product_ID	Product_Line	Product_Category	Quantity	Retail_Price	Cost_Price	Customer_Country	Customer_Continent	Customer_Dob
4	1244336156	31DEC2018	31DEC2018	1	240500000000	Sports	Running - Jogging	-1	117.8	52.2	IT	Europe	1995-04-23
99	1244319311	29DEC2018	29DEC2018	1	210201000000	Children	Children Sports	-1	55.98	48.1	BE	Europe	1965-08-01
168	1244302076	27DEC2018	27DEC2018	1	230100000000	Outdoors	Outdoors	-1	283.2	119.2	FR	Europe	1985-04-03

/* Found : 3 Records with anomaly */

```
/*=============================================================*/
                /*===== Cleaning the data =====*/
/*=============================================================*/

/*=============================================================*/
/* Delivery Date Issue Fix*/
/*=============================================================*/

data orders_clean1;
set orders;
where delivery_date >= order_date;
run;

/*=============================================================*/
/* Verifying if the delivery date issue was fixed */
/*=============================================================*/

proc print data=orders_clean1;
where order_date > delivery_date;
run;
```

```
1          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
68
69         proc print data=orders_clean1;
70         where order_date > delivery_date;
71         run;

NOTE: No observations were selected from data set WORK.ORDERS_CLEAN1.
NOTE: There were 0 observations read from the data set WORK.ORDERS_CLEAN1.
      WHERE order_date>delivery_date;
```

```
        /* No records found, issue was fixed. */


/*=============================================================*/
/* Invalid Order Type Issue Fix */
/*=============================================================*/

data orders_clean2;
set orders_clean1;
length Order_Type_Detail $10;
select(order_type);
when(1) Order_Type_Detail = "Retail";
when(2) Order_Type_Detail = "Phone";
when(3) Order_Type_Detail = "Internet";
otherwise Order_Type_Detail = "Unknown";
end;
run;
```

```
/*============================================================*/
/* Verifying if the invalid order type issue was fixed */
/*============================================================*/

proc freq data=orders_clean2;
table Order_Type_Detail;
```

Order_Type_Detail	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Internet	1356	12.58	1356	12.58
Phone	1089	10.10	2445	22.68
Retail	8333	77.29	10778	99.97
Unknown	3	0.03	10781	100.00

```
   /* New column was successfully added with defined classes. */


/*============================================================*/
/* Column Quantity With Negative/<0 Values Issue */
/*============================================================*/

data orders_clean4;
set orders_clean3;
if quantity < 0 then quantity = 0;
run;

/*============================================================*/
/* Verifying if quantity with negative values issue was fixed */
/*============================================================*/

proc means data=orders_clean4;
var quantity;
run;
```

Analysis Variable : Quantity				
N	Mean	Std Dev	Minimum	Maximum
10781	1.6807346	0.8959344	0	8.0000000

```
/* Issue was fixed successfully, minimum value for quantity
variable is now ‘0’. */
```

```
/*=============================================================*/
/* Invalid Case of Country Key Issue Fix */
/*=============================================================*/

data orders_clean3;
set orders_clean2;
customer_country = upcase(customer_country);
run;

/*=============================================================*/
/* Verifying if the invalid case of country key issue was fixed */
/*=============================================================*/
proc freq data=orders_clean3;
table(customer_country);
```

Customer_Country	Frequency	Percent	Cumulative Frequency	Cumulative Percent
AE	2	0.02	2	0.02
AT	12	0.11	14	0.13
AU	732	6.79	746	6.92
BE	418	3.88	1164	10.80
CA	44	0.41	1208	11.20
CH	13	0.12	1221	11.33
CZ	1	0.01	1222	11.33
DE	1282	11.89	2504	23.23
DK	182	1.69	2686	24.91
ES	1056	9.80	3742	34.71
FI	12	0.11	3754	34.82
FR	1280	11.87	5034	46.69
GB	1213	11.25	6247	57.94
GR	2	0.02	6249	57.96
HR	1	0.01	6250	57.97
HU	1	0.01	6251	57.98
IE	2	0.02	6253	58.00
IL	1	0.01	6254	58.01
IT	1383	12.83	7637	70.84
LU	3	0.03	7640	70.87
NL	702	6.51	8342	77.38
NO	8	0.07	8350	77.45
PT	12	0.11	8362	77.56
RU	1	0.01	8363	77.57
SE	9	0.08	8372	77.66
SI	2	0.02	8374	77.67
TR	8	0.07	8382	77.75
US	2385	22.12	10767	99.87
ZA	14	0.13	10781	100.00

```
/* Issue was fixed, successfully, all country keys with lowercase
were converted to upper case and merged with the originals. */
```

```
/*================================================================*/
      /* ==== PRE-PROCESSING THE DATA FOR ANALYSIS =====*/
/*================================================================*/

/* Creating custom-format agegrp (age-group) for customer-age */
proc format;
value agegrp
LOW-10 = 'Kids'
11-17 = 'Teenagers'
18-24 = 'College Students'
25-55 = 'Working Professionals'
55-65 = 'Retired'
65-HIGH = 'Elders'
;

/*================================================================*/
/*
Deriving following columns for analysis:
1. Profit - To calculate profit per order
2. Shipping Days - To calculate total shipping days per order
3. Age - To calculate customer's age (from customer_dob)
4. Age-Group - Using Age to classify customer for a particular
age-group
*/
/*================================================================*/

data orders_clean5
(drop = age customer_dob retail_price cost_price quantity);

set orders_clean4;

Age = intck('year',customer_dob,today());

Age_Group = put(age,agegrp.);

Profit = (retail_price-cost_price)*quantity;

Shipping_Days = intck('day',order_date,delivery_date);
r
un;
```

```
/*=============================================================*/
          /*==== Combining Data-Sets === */
/*=============================================================*/

/*=============================================================*/
/* Sorting country_clean data by country_key */
/*=============================================================*/

proc sort data=country_clean;
by country_key;

/*=============================================================*/
/* Sorting cleaned orders data by customer_country key */
/*=============================================================*/

proc sort data=orders_clean5;
by customer_country;

/*=============================================================*/
/* Combining Cleaned Orders Data with Country-Population Data */
/*=============================================================*/

data final_clean(drop = lat lon);
merge country_clean(in=in1 rename=(country_key=customer_country))
orders_clean5(in=in2);
by customer_country;
if in1=1 and in2=1;
run;

/*=============================================================*/
/* Viewing first 10 records from final-cleaned-data-set */
/*=============================================================*/

proc print data=final_clean (obs=10);
run;
```

Obs	customer_country	Country_Name	Order_ID	Order_Date	Delivery_Date	Order_Type	Product_ID	Product_Line	Product_Category	Customer_Continent	Order_Type_Detail	Age_Group	Profit	Shipping_Days
1	AE	United Arab Emirates	1244217525	18DEC2018	19DEC2018	3	240100000000	Sports	Assorted Sports Articles	Asia	Internet	College Students	649.40	1
2	AE	United Arab Emirates	1241995637	17APR2018	18APR2018	3	240100000000	Sports	Assorted Sports Articles	Asia	Internet	College Students	104.20	1
3	AT	Austria	1243127534	20AUG2018	21AUG2018	3	240100000000	Sports	Assorted Sports Articles	Europe	Internet	Retired	361.40	1
4	AT	Austria	1243049921	11AUG2018	17AUG2018	3	240100000000	Sports	Assorted Sports Articles	Europe	Internet	Working Professionals	265.95	6
5	AT	Austria	1242396370	31MAY2018	04JUN2018	2	240100000000	Sports	Assorted Sports Articles	Europe	Phone	Working Professionals	119.25	4
6	AT	Austria	1242223869	12MAY2018	14MAY2018	3	240100000000	Sports	Assorted Sports Articles	Europe	Internet	Working Professionals	127.85	2
7	AT	Austria	1242060255	24APR2018	30APR2018	3	240100000000	Sports	Assorted Sports Articles	Europe	Internet	Working Professionals	9.30	6
8	AT	Austria	1241908505	07APR2018	12APR2018	3	240100000000	Sports	Assorted Sports Articles	Europe	Internet	Working Professionals	4.95	5
9	AT	Austria	1241908505	07APR2018	12APR2018	3	240100000000	Sports	Assorted Sports Articles	Europe	Internet	Working Professionals	11.15	5
10	AT	Austria	1241897622	06APR2018	19APR2018	3	240100000000	Sports	Assorted Sports Articles	Europe	Internet	Working Professionals	638.40	13

** Final-Dataset was created successfully **

```
/*================================================================*/
       /* ====== PERFORMING SALES ANALYSIS ====== */
/*================================================================*/

/* Creating excel file to export sales analysis output to excel
sheets */

ods excel file="&path.\sales_output.xlsx";

/*================================================================*/
       /* ===== Orders Frequency Analysis ===== */
/*================================================================*/

/*================================================================*/
* Which months have highest and lowest total number of orders? *
/*================================================================*/

/*================================================================*/
ANSWER:

Highest Number Of Orders =>
Month : December
Orders : 1200
%of Total Orders : 11.27

Lowest Number Of Orders =>
Month : February
Orders : 700
%of Total Orders : 6.49

Check Output: */
/*================================================================*/

proc freq data=final_clean order=freq;
table order_date/nocum;
format order_date monname.;
```

Order_Date	Frequency	Percent
December	1200	11.13
August	1026	9.52
November	984	9.13
July	981	9.10
May	964	8.94
June	913	8.47
October	908	8.42
April	886	8.22
January	750	6.96
September	741	6.87
March	728	6.75
February	700	6.49

```
/*============================================================*/
 * How many orders are distributed by each continent? *
/*============================================================*/

/*============================================================*/
ANSWER:

Highest Number Of Orders =>
Continent : Europe
Orders : 7594
%of Total Orders : 70.44

Lowest Number Of Orders =>
Continent : Asia
Orders : 12
%of Total Orders : 0.11

Check Output:
/*============================================================*/

proc freq data=final_clean order=freq;
table customer_continent/nocum;
```

Customer_Continent	Frequency	Percent
Europe	7594	70.44
North	2429	22.53
Austra	732	6.79
Africa	14	0.13
Asia	12	0.11

```
/*============================================================*/
Within each continent how many orders were placed via retail,
internet or phone?
/*============================================================*/

/*============================================================*/
ANSWER:
Check Output:
/*============================================================*/

proc freq data=final_clean order=freq;
table customer_continent*Order_Type_Detail/norow nocol nopercent;
```

Table of Customer_Continent by Order_Type_Detail

Customer_Continent	Order_Type_Detail				
	Retail	Internet	Phone	Unknown	Total
Europe	6002	848	741	3	7594
North	1772	384	273	0	2429
Austra	559	106	67	0	732
Africa	0	9	5	0	14
Asia	0	9	3	0	12
Total	8333	1356	1089	3	10781

```
/*============================================================*/
          /* ===== Ship Days Summary =====*/
/*============================================================*/
* How many days on average does it takes for an order to be
delivered? *
/*============================================================*/


/*============================================================*/
ANSWER:

On an Average it takes 3-4 days for an order to be delivered.

Check Output:
/*============================================================*/

proc means data=final_clean mean maxdec=2;
var shipping_days;
where shipping_days > 0;
```

Analysis Variable : Shipping_Days
Mean
3.71

```
/*============================================================*/
Are there any countries where shipment takes longer?
/*============================================================*/

/*============================================================*/
ANSWER:
/*============================================================*/

Countries like Austria, Norway, Switzerland are places where
shipment takes longer than average (i.e. >3-4 days).

Check Output:
/*============================================================*/

proc means data=final_clean mean maxdec=2 ;
var shipping_days;
class country_name;
where shipping_days > 0;
```

Analysis Variable : Shipping_Days		
Country_Name	N Obs	Mean
Australia	208	3.89
Austria	12	4.92
Belgium	86	3.42
Canada	44	3.27
Croatia	1	3.00
Czech Republic	1	1.00
Denmark	51	3.47
Finland	12	2.83
France	329	3.74
Germany	326	3.72
Greece	2	3.50
Hungary	1	2.00
Ireland	2	2.50
Israel	1	4.00
Italy	337	3.61
Luxembourg	2	1.50
Netherlands	195	3.47
Norway	8	4.50
Portugal	12	4.25
Russia	1	4.00
Slovenia	2	2.50
South Africa	14	3.36
Spain	302	3.76
Sweden	9	2.78
Switzerland	13	4.00
Turkey	8	3.00
United Arab Emirates	2	1.00
United Kingdom	235	4.07
United States	737	3.71

```
/*============================================================*/
     /* ===== Profit Analysis by Customer Age ===== */
/*============================================================*/

/*============================================================*/
Which customer age group produces the highest median profit per
order ?
/*============================================================*/

/*============================================================*/
ANSWER:

The age-group 'College-Students' produces highest (62.35) median
profit per order.

The other two age-groups that come close to making similar median
profit per order are 'Teenagers' (61.83) and 'Working
Professionals' (60.55).

Check Output:
/*============================================================*/

proc means data=final_clean median maxdec=2 nway;
var profit;
class age_group;
```

Analysis Variable : Profit		
Age_Group	N Obs	Median
College Students	1617	62.35
Retired	1389	59.20
Teenagers	462	61.83
Working Professionals	7313	60.55

```
* Closing excel file *

ods excel close

/* NOTE: All the outputs for questions have been exported in excel
file titled ‘sales_output.xlsx’ stored in project folder.


/*=====================---END---==============================*/
```