AWESOME CHOCOLATES DATA CLEANING

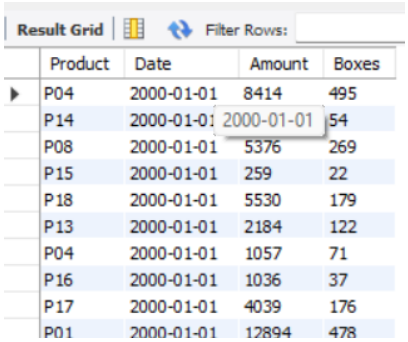
BY USING SQL

This project focuses on cleaning and analysing sales data to derive actionable insights using SQL. The process involves several key steps to ensure data quality and integrity

By executing these SQL queries, the project ensures a robust data cleaning process, leading to accurate and insightful analysis. The cleaned and enriched data is then ready for further exploration and visualization, ultimately aiding in data-driven decision-making and strategy formulation.

1 – See all shipments

```
select * from shipment;

select Product, s.Date, Amount, Boxes from shipment s;
```

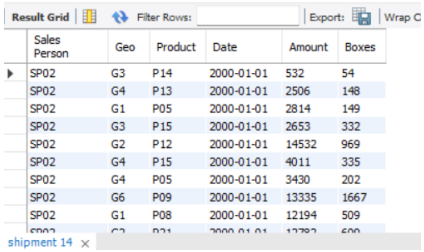


Result Grid | Filter Rows:

Product	Date	Amount	Boxes
P04	2000-01-01	8414	495
P14	2000-01-01	2000-01-01	54
P08	2000-01-01	5376	269
P15	2000-01-01	259	22
P18	2000-01-01	5530	179
P13	2000-01-01	2184	122
P04	2000-01-01	1057	71
P16	2000-01-01	1036	37
P17	2000-01-01	4039	176
P01	2000-01-01	12894	478

2 – All shipments by SP02

```
select * from shipment s

where s.`Sales Person` ='SP02';
```



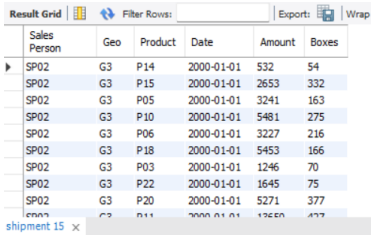
Result Grid | Filter Rows: | Export: | Wrap C

Sales Person	Geo	Product	Date	Amount	Boxes
SP02	G3	P14	2000-01-01	532	54
SP02	G4	P13	2000-01-01	2506	148
SP02	G1	P05	2000-01-01	2814	149
SP02	G3	P15	2000-01-01	2653	332
SP02	G2	P12	2000-01-01	14532	969
SP02	G4	P15	2000-01-01	4011	335
SP02	G4	P05	2000-01-01	3430	202
SP02	G6	P09	2000-01-01	13335	1667
SP02	G1	P08	2000-01-01	12194	509

shipment 14 ×

3 – All shipments by SP02 to G3

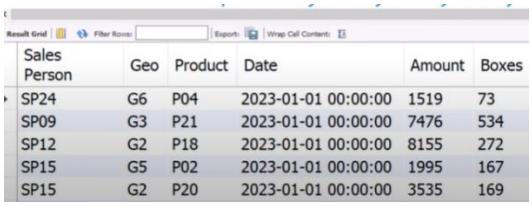
```
select * from shipment s

where s.`Sales Person` ='SP02' and s.`Geo` = 'G3';
```



Sales Person	Geo	Product	Date	Amount	Boxes
SP02	G3	P14	2000-01-01	532	54
SP02	G3	P15	2000-01-01	2653	332
SP02	G3	P05	2000-01-01	3241	163
SP02	G3	P10	2000-01-01	5481	275
SP02	G3	P06	2000-01-01	3227	216
SP02	G3	P18	2000-01-01	5453	166
SP02	G3	P03	2000-01-01	1246	70
SP02	G3	P22	2000-01-01	1645	75
SP02	G3	P20	2000-01-01	5271	377

4 – All shipments in Jan 2023

```
select * from shipment s

where s. Date between '2023-01-01' and '2023-01-31';
```



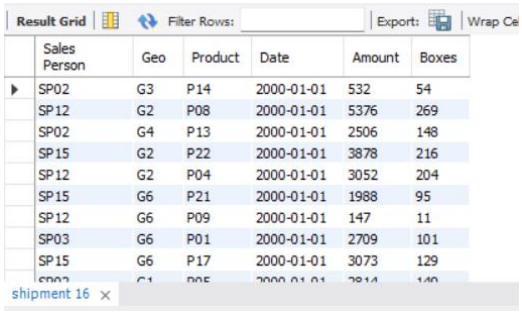
Sales Person	Geo	Product	Date	Amount	Boxes
SP24	G6	P04	2023-01-01 00:00:00	1519	73
SP09	G3	P21	2023-01-01 00:00:00	7476	534
SP12	G2	P18	2023-01-01 00:00:00	8155	272
SP15	G5	P02	2023-01-01 00:00:00	1995	167
SP15	G2	P20	2023-01-01 00:00:00	3535	169

5 – All shipments by SP02, SP03, SP12, SP15

```
select * from shipment s

where s.`Sales Person` ='SP02'

      or s.`Sales Person` ='SP03'

      or s.`Sales Person` ='SP12'

      or s.`Sales Person` ='SP15'
```

OR

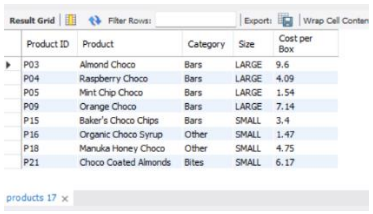
```
select * from shipment s

where s.`Sales Person` in ('SP03', 'SP12', 'SP15' );
```



Sales Person	Geo	Product	Date	Amount	Boxes
SP02	G3	P14	2000-01-01	532	54
SP12	G2	P08	2000-01-01	5376	269
SP02	G4	P13	2000-01-01	2506	148
SP15	G2	P22	2000-01-01	3878	216
SP12	G2	P04	2000-01-01	3052	204
SP15	G6	P21	2000-01-01	1988	95
SP12	G6	P09	2000-01-01	147	11
SP03	G6	P01	2000-01-01	2709	101
SP15	G6	P17	2000-01-01	3073	129

6 – Products that have the word choco in them

```
select * from products

where product like '%choco%';
```

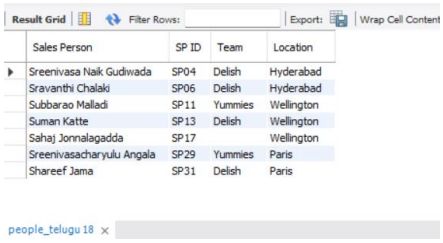


Product ID	Product	Category	Size	Cost per Box
P03	Almond Choco	Bars	LARGE	9.6
P04	Raspberry Choco	Bars	LARGE	4.09
P05	Mint Chip Choco	Bars	LARGE	1.54
P09	Orange Choco	Bars	LARGE	7.14
P15	Baker's Choco Chips	Bars	SMALL	3.4
P16	Organic Choco Syrup	Other	SMALL	1.47
P18	Manuka Honey Choco	Other	SMALL	4.75
P21	Choco Coated Almonds	Bites	SMALL	6.17

products 17

7 – Sales persons whose name begins with S

```
select * from people_telugu

where `sales Person` like 'S%';
```

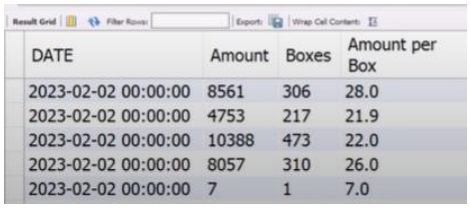


Sales Person	SP ID	Team	Location
Sreenivasa Naik Gudiwada	SP04	Delish	Hyderabad
Sravanthi Chalaki	SP06	Delish	Hyderabad
Subbarao Malladi	SP11	Yummies	Wellington
Suman Katte	SP13	Delish	Wellington
Sahaj Jonnalagadda	SP17		Wellington
Sreenivasacharyulu Angala	SP29	Yummies	Paris
Shareef Jama	SP31	Delish	Paris

people_telugu 18

8 – Sales per box of chocolates in Feb 2023

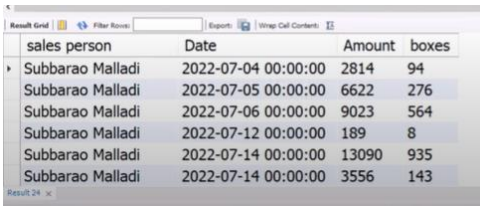
```
SELECT S.DATE, s.Amount, s.Boxes, round(S.Amount / S.Boxes, 1) as 'Amount per Box' FROM
SHIPMENTS_NEW S
WHERE extract(year_month from s.Date) = 202302;
```



DATE	Amount	Boxes	Amount per Box
2023-02-02 00:00:00	8561	306	28.0
2023-02-02 00:00:00	4753	217	21.9
2023-02-02 00:00:00	10388	473	22.0
2023-02-02 00:00:00	8057	310	26.0
2023-02-02 00:00:00	7	1	7.0

9 – All shipment data for Subbarao

```
select * from people
where `sales person` like 'subba%';

select * from shipments_new
where `sales person` = 'sp11';

select p.`sales person`, s.Date, s.Amount, s.boxes from shipments_new s
join people p on p.`sp id` = s.`sales person`
where p.`sales person` LIKE 'Subba%';
```



sales person	Date	Amount	boxes
Subbarao Malladi	2022-07-04 00:00:00	2814	94
Subbarao Malladi	2022-07-05 00:00:00	6622	276
Subbarao Malladi	2022-07-06 00:00:00	9023	564
Subbarao Malladi	2022-07-12 00:00:00	189	8
Subbarao Malladi	2022-07-14 00:00:00	13090	935
Subbarao Malladi	2022-07-14 00:00:00	3556	143

Result 24

10 – All shipment data for Subbarao by month

```
select extract(year_month from s.Date), sum(s.Amount), sum(s.boxes)
from shipments_new s
```

```
join people p on p.`sp id` = s.`sales person`
where p.`sales person` LIKE 'Subba%'
group by extract(year_month from s.Date);
```

extract(year_month from s.Date)	sum(s.Amount)	sum(s.boxes)
202207	152159	8909
202208	88963	4933
202209	149870	10803
202210	106281	7188
202211	91112	5511
202212	116998	8894
202301	124348	7159
202302	66696	4784

11 – All shipment data for Subbarao to USA

```
select p.`sales person`, g.Geo, s.Amount, s.boxes from shipments_new s
join people p on p.`sp id` = s.`sales person`
join geo g on g.GeoID = s.Geo
where
p.`sales person` LIKE 'Subba%' and
g.Geo = "USA";
```

#12 – What is the maximum amount in each month?

```
select extract(year_month from s.Date), max(s.Amount), min(s.amount)
from shipments_new s
group by extract(year_month from s.Date);
```

13 – How many shipments we do by each country in the month of March 2023

```
select g.geo, count(*), sum(s.Amount)
from shipments_new s
join geo g on g.GeoID = s.Geo
where extract(year_month from s.Date) = 202303
group by g.Geo;
```