

Walmart Data Analysis using MySQL

3 • `use walmart_data;`

Action Output				
#	Time	Action	Message	Duration / Fetch
✓ 1	15:25:47	use walmart_data	0 row(s) affected	0.000 sec

The data has been imported through table import wizard and the data type has been check before importing.

4 • `select * from walmart;`

5

Result Grid										
Filter Rows:										
Export: Wrap Cell Content:										
	Invoice ID	Branch	City	Customer type	Gender	Product line	Unit price	Quantity	Tax 5%	Total
▶	750-67-8428	A	Yangon	Member	Female	Health and beauty	74.69	7	26.1415	548.97
	226-31-3081	C	Naypyitaw	Normal	Female	Electronic accessories	15.28	5	3.82	80.22
	631-41-3108	A	Yangon	Normal	Male	Home and lifestyle	46.33	7	16.2155	340.52
	123-19-1176	A	Yangon	Member	Male	Health and beauty	58.22	8	23.288	489.04
	373-73-7910	A	Yangon	Normal	Male	Sports and travel	86.31	7	30.2085	634.37
	699-14-3026	C	Naypyitaw	Normal	Male	Electronic accessories	85.39	7	29.8865	627.61

walmart 1 x Read Only

Output

Action Output				
#	Time	Action	Message	Duration / Fetch
✓ 1	15:25:47	use walmart_data	0 row(s) affected	0.000 sec
✓ 2	15:27:06	select * from walmart	1000 row(s) returned	0.000 sec / 0.000 sec

6 • `select count(*) `Total Count` from walmart;`

Result Grid		Filter Rows:	
	Total Count		
▶	1000		

#	Time	Action	Message	Duration / Fetch
✓ 2	15:27:06	select * from walmart	1000 row(s) returned	0.000 sec / 0.000 sec
✓ 3	15:28:33	select count(*) `Total Count` from walmart	1 row(s) returned	0.000 sec / 0.000 sec

8 `-- 1.Generic Question`

9 `-- 1.a) How many unique cities does the data have?`

10 • `select count(distinct(city)) CityCount from walmart;`

Result Grid		Filter Rows:	
	CityCount		
▶	3		

Action Output				
#	Time	Action	Message	Duration / Fetch
✓ 3	15:28:33	select count(*) `Total Count` from walmart	1 row(s) returned	0.000 sec / 0.000 sec
✓ 4	15:30:08	select count(distinct(city)) CityCount from walmart	1 row(s) returned	0.000 sec / 0.000 sec

```

12  -- 1.b) In which city, each branch is located?
13  •  select distinct(branch),city
14      from walmart
15      order by branch,city;



```

Result Grid		Filter Rows:
	branch	city
▶	A	Yangon
	B	Mandalay
	C	Naypyitaw

```

17  -- 2.Product
18  -- 2.a) How many unique product lines does the data have?
19  •  select count(distinct(`product line`)) UniqueProductLine
20      from walmart;

```


Result Grid   Filter Rows:

	UniqueProductLine
▶	6

```

22  -- 2.b) What is the most common payment method?
23  •  select max(payment) `Most Common Payment`
24      from walmart;
25

```

Result Grid			Filter Rows:
	Most Common Payment		
▶	Ewallet		

```

26  -- 2.c) What is the most selling product line?
27  •  select max(`product line`) `Most Selling ProductLine`
28      from walmart;

```

	Most Selling ProductLine
▶	Sports and travel

```

30  -- 2.d) What is the total revenue by month?
31  • select monthname(date) Month_Name,round(sum(total),2) total_revenue
32  from walmart
33  group by month_name
34  order by total_revenue desc;

```

Result Grid			Filter Rows:
	Month_Name	total_revenue	
▶	January	116291.87	
	March	109455.51	
	February	97219.37	

```

36  -- 2.e) Which month had the largest COGS?
37  • select monthname(date) month_name,round(sum(cogs),2) max_cogs
38  from walmart
39  group by month_name
40  order by max_cogs desc limit 1;
41

```

Result Grid			Filter Rows:	Export:	Wrap Cell Content:
	month_name	max_cogs			
▶	January	110754.16			

```

42  -- 2.f) What product line had the largest revenue?
43  • select `product line`, round(sum(total),2) total_revenue
44  from walmart
45  group by `product line`
46  order by total_revenue desc limit 1;
47

```

Result Grid			Filter Rows:	Export:	Wrap Cell Content:
	product line	total_revenue			
▶	Food and beverages	56144.84			

```

48  -- 2.g) What is the city with the largest revenue?
49  • select city,branch,round(sum(total),2) total_revenue
50  from walmart
51  group by city,branch
52  order by total_revenue desc limit 1;
53

```

Result Grid				Filter Rows:	Export:	Wrap Cell Content:
	city	branch	total_revenue			
▶	Naypyitaw	C	110568.71			

```

54 -- 2.h) What product line had the largest VAT?
55 • select `product line`, round(avg(`tax 5%`),2) tax
56 from walmart
57 group by `product line`
58 order by tax desc limit 1;
59

```

Result Grid		Filter Rows:	Export:	Wrap Cell Cor
	product line	tax		
▶	Home and lifestyle	16.03		

```

61 -- 2.i) Fetch each product line and add a column to those product line showing "Good","Bad".
62 -- Good if its greater than average sales and find the invoice ID which are "Good"
63 -- the first query is done to find out the average number of products from each productline.
64 • select `product line`,avg(quantity)
65 from walmart
66 group by `product line`;
67
68 • select `invoice id`,`product line`, quantity,
69 case
70 when quantity >= 6 then "Good"
71 when quantity < 6 then "Bad"
72 end SalesQuality
73 from walmart;

```

Result Grid		Filter Rows:	Export:	Wrap Cell Content:
	invoice id	product line	quantity	SalesQuality
▶	750-67-8428	Health and beauty	7	Good
	226-31-3081	Electronic accessories	5	Bad
	631-41-3108	Home and lifestyle	7	Good

```

74 -- to retrieve the SalesQuality Column containing only Good
75 • with CTE as
76 (
77 select `invoice id`,`product line`, quantity,
78 case
79 when quantity >= 6 then "Good"
80 when quantity < 6 then "Bad"
81 end SalesQuality
82 from walmart
83 )
84 select `invoice id`,`product line`
85 from CTE
86 where SalesQuality = "Good";

```

Result Grid		Filter Rows:	Export:	Wrap Cell Content:
	invoice id	product line		
▶	750-67-8428	Health and beauty		
	631-41-3108	Home and lifestyle		
	123-19-1176	Health and beauty		
	373-73-7910	Sports and travel		

```

89 -- 2.j)The average number of products sold at which branch
90 -- is higher than that of the total average number of products sold
91 • select branch,round(avg(quantity),2) AvgSales
92 from walmart
93 group by branch
94 having AvgSales > (select round(avg(quantity),2) TotalAvgSales
95 from walmart);
96

```

Result Grid		Filter Rows:	Export:	Wrap Cell Content:
branch	AvgSales			
▶ C	5.58			

```

97 -- 2.k)What is the most common product line by gender?
98 • select `product line`,gender ,count(gender) as total_count
99 from walmart
100 group by `product line`,gender
101 order by total_count desc,gender asc;
102

```

Result Grid		Filter Rows:	Export:	Wrap Cell Content:
product line	gender	total_count		
▶ Fashion accessories	Female	96		
Food and beverages	Female	90		
Sports and travel	Female	88		
Health and beauty	Male	88		
Electronic accessories	Male	86		

```

103 -- 2.l) What is the average rating of each product line?
104 • select `product line`,round(avg(rating),2) AvgRating
105 from walmart
106 group by `product line`
107 order by AvgRating desc;
108

```

Result Grid		Filter Rows:	Export:	Wrap Cell Content:
product line	AvgRating			
▶ Food and beverages	7.11			
Fashion accessories	7.03			
Health and beauty	7			
Electronic accessories	6.92			
Sports and travel	6.92			

```

109 -- 2.m) Calculate the total sales (total) for each product line (product_line).
110 • select `product line`,round(sum(total),2) TotalSales
111 from walmart
112 group by `product line`
113 order by TotalSales desc;
114

```

Result Grid	Filter Rows:	Export:	Wrap Cell Content:
product line	TotalSales		
Food and beverages	56144.84		
Sports and travel	55122.83		
Electronic accessories	54337.53		
Fashion accessories	54305.9		
Home and lifestyle	53861.91		

```

115 -- 3. Branch Performance Analysis:
116 -- 3. a) Find the total sales (total) for each branch.
117 • select branch, round(sum(total),2) TotalSales
118 from walmart
119 group by branch;
120

```

Result Grid	Filter Rows:	Export:	Wrap Cell Content:
branch	TotalSales		
A	106200.37		
C	110568.71		
B	106197.67		

```

121 -- 3. b) Calculate the average rating (rating) for each branch.
122 • select branch, round(avg(rating),2) AvgRating
123 from walmart
124 group by branch;

```

Result Grid	Filter Rows:	Export:	Wrap Cell Content:
branch	AvgRating		
A	7.03		
C	7.07		
B	6.82		

```

126 -- 4. Customer Segmentation:
127 -- 4.a) Identify the number of customers (invoice_id) in each city (city).
128 • select city,count(`invoice id`) NoOfCustomers
129 from walmart
130 group by city;

```

Result Grid	Filter Rows:	Export:	Wrap Cell Content:
city	NoOfCustomers		
Yangon	340		
Naypyitaw	328		
Mandalay	332		


```

132 -- 4.b) Determine the total sales (total) for each customer type (customer_type).
133 • select `customer type`,round(sum(total),2) TotalSales
134 from walmart
135 group by `customer type`;

```

Result Grid			Filter Rows:	Export:	Wrap Cell Content:
	customer type	TotalSales			
▶	Member	164223.44			
	Normal	158743.31			

```

137 -- 4.c) How many unique customer types does the data have?
138 • select `customer type`,count(`customer type`) TotalCount
139 from walmart
140 group by `customer type`;

```

Result Grid			Filter Rows:	Export:	Wrap Cell Content:
	customer type	TotalCount			
▶	Member	501			
	Normal	499			

```

143 -- 4.d) How many unique payment methods does the data have?
144 • select payment, count(payment) TotalCount
145 from walmart
146 group by payment;

```

Result Grid			Filter Rows:	Export:	Wrap Cell Content:
	payment	TotalCount			
▶	Ewallet	345			
	Cash	344			
	Credit card	311			

```

148 -- 4.e) Which customer type buys the most(by quantity)?
149 • select *,
150 dense_rank() over(order by TotalQuantity desc) rn
151 from ( select `customer type`, sum(quantity) TotalQuantity
152 from walmart
153 group by `customer type`
154 ) temp;

```

Result Grid				Filter Rows:	Export:	Wrap Cell Content:
	customer type	TotalQuantity	rn			
▶	Member	2785	1			
	Normal	2725	2			

```

156 -- 4.f) What is the gender of most of the customers?
157 • select gender, count(gender) TotalCount
158 from walmart
159 group by gender;

```

Result Grid	Filter Rows:	Export:	Wrap Cell Content
gender	TotalCount		
Female	501		
Male	499		

```

161 -- 4.g) What is the gender distribution per branch?
162 • select branch, gender, count(gender) GenderCount
163 from walmart
164 group by branch, gender
165 order by branch, GenderCount desc;

```

Result Grid	Filter Rows:	Export:	Wrap Cell Content
branch	gender	GenderCount	
A	Male	179	
A	Female	161	
B	Male	170	
B	Female	162	
C	Female	178	

```

167 -- 4.h) Which time of the day do customers give most ratings?
168 • alter table walmart add column part_of_day varchar(20);
169
170 • set sql_safe_updates = 0;
171
172 • update walmart
173   set part_of_day = case
174     when time between "00:00:00" and "11:59:59" then "Morning"
175     when time between "12:00:00" and "15:59:59" then "Afternoon"
176     else "Evening"
177   end;
178 • select * from walmart;
179
180 • select part_of_day, round(avg(rating),2) AvgRating
181 from walmart
182 group by part_of_day;

```

Result Grid	Filter Rows:	Export:	Wrap Cell Content:
part_of_day	AvgRating		
Afternoon	7.03		
Morning	6.96		
Evening	6.93		


```

184 -- 4.i) Which time of the day do customers give most ratings per branch?
185 • select branch,part_of_day,round(avg(rating),2) AvgRating
186 from walmart
187 group by branch,part_of_day
188 order by AvgRating desc;

```

Result Grid			
		Filter Rows:	
		Export:	
		Wrap Cell Content:	
	branch	part_of_day	AvgRating
▶	A	Afternoon	7.19
	C	Evening	7.12
	C	Afternoon	7.07
	A	Morning	7.01

```

190 -- 4.j) Which day for the week has the best avg ratings?
191 • select dayname(date)
192 from walmart;
193
194 • alter table walmart add column dayname varchar(10);
195
196 • update walmart
197 set `dayname` = dayname(date);
198 • select * from walmart;
199
200 • select dayname,round(avg(rating),2) AvgRating
201 from walmart
202 group by dayname
203 order by AvgRating desc;

```

Result Grid		
		Filter Rows:
		Export:
		Wrap Cell Content:
	dayname	AvgRating
▶	Monday	7.15
	Friday	7.08
	Sunday	7.01

```

205 -- 4.k) Which day of the week has the best average ratings per branch?
206 • select branch,dayname,round(avg(rating),2) AvgRating
207 from walmart
208 group by branch,dayname
209 order by AvgRating desc;

```

Result Grid		
		Filter Rows:
		Export:
		Wrap Cell Content:
	dayname	AvgRating
▶	Monday	7.15
	Friday	7.08
	Sunday	7.01
	Tuesday	7
	Saturday	6.9

```

211 -- 5. Sales Trends:
212 -- 5.a) Determine the total sales (total) for each year.
213 • select year(date)Year,round(sum(total),2) TotalSales
214 from walmart
215 group by Year;

```

Result Grid   Filter Rows: Export:  Wrap Cell Content:

	Year	TotalSales
▶	2019	322966.75

```

217 -- 5.b) Find the month with the highest total sales (total) across all years.
218 • select monthname(date) Month,round(sum(total),2) TotalSales
219 from walmart
220 group by Month
221 order by TotalSales desc;

```


Result Grid   Filter Rows: Export:  Wrap Cell Content: 

	Month	TotalSales
▶	January	116291.87
	March	109455.51
	February	97219.37

```

223 -- 5.c) Number of sales made in each time of the day per weekday
224 • with cte as
225 (
226 select month(date) Month,weekday(date) weekday,hour(time) hour,count(`invoice ID`) TotalCount
227 from walmart
228 group by Month, weekday, hour
229 order by month asc
230 )
231 select case
232 when month = 1 then "January"
233 when month = 2 then "February"
234 when month = 3 then "March"
235 end MonthName, weekday,hour, totalcount
236 from cte;

```





Result Grid  Filter Rows: Export:  Wrap Cell Content: 

	MonthName	weekday	hour	totalcount
▶	January	0	10	5
	January	0	11	2
	January	0	12	7
	January	0	13	2

```

238 -- 5.d) Which of the customer types brings the most revenue?
239 • select `customer type`,round(sum(total),2) TotalRevenue
240 from walmart
241 group by `customer type`
242 order by TotalRevenue desc;

```

Result Grid   Filter Rows: Export:  Wrap Cell Content: 

	customer type	TotalRevenue
▶	Member	164223.44
	Normal	158743.31

```

244 -- 6. Payment Method Analysis:
245 -- 6.a) Calculate the average payment amount (payment_method)
246 -- for each payment method (payment_method).
247 • select payment,round(avg(total),2) AvgAmountSpent
248 from walmart
249 group by payment
250 order by AvgAmountSpent desc;

```

Result Grid	Filter Rows:	Export:	Wrap Cell Content:
payment	AvgAmountSpent		
Cash	326.18		
Credit card	324.01		
Ewallet	318.82		

```

252 -- 6.b) Identify the most common payment method (payment_method) used by customers.
253 • select payment,count(payment) TotalCount
254 from walmart
255 group by payment
256 order by TotalCount desc;

```

Result Grid	Filter Rows:	Export:	Wrap Cell Content:
payment	TotalCount		
Ewallet	345		
Cash	344		
Credit card	311		

```

258 -- 7. Time Analysis:
259 -- 7.a) Analyze the total sales (total) made during each hour of the day (time).
260 • select hour(time) Hour,round(sum(total),2) TotalSales
261 from walmart
262 group by Hour
263 order by hour asc;

```

Result Grid	Filter Rows:	Export:	Wrap Cell Content:
Hour	TotalSales		
10	31421.48		
11	30377.33		
12	26065.88		
13	34723.23		
14	30828.4		

```

265 -- 7.b) Determine the day of the week with the highest average sales (total)
266 • select dayname,round(sum(total),2) TotalSales,round(avg(total),2) AvgSales
267 from walmart
268 group by dayname
269 order by AvgSales desc;

```

Result Grid	Filter Rows:	Export:	Wrap Cell Content:
dayname	TotalSales	AvgSales	
Saturday	56120.81	342.2	
Sunday	44457.89	334.27	
Thursday	45349.25	328.62	
Tuesday	51482.25	325.84	
Friday	43926.34	316.02	

```

271 -- 8. Customer Satisfaction Analysis:
272 -- 8.a) Find the average rating (rating) for each gender (gender).
273 • select gender,round(avg(rating),2) AvgRating
274 from walmart
275 group by gender
276 order by AvgRating desc;

```

Result Grid		
	gender	AvgRating
▶	Male	6.98
	Female	6.96

```

278 -- 8.b) Determine the gender (gender) with the highest average rating (rating)
279 • with cte as
280 (
281 select gender,rating,
282 dense_rank() over(order by rating desc) rnk
283 from walmart
284 )
285 select gender,rating
286 from cte
287 where rnk = 1;

```

Result Grid		
	gender	rating
▶	Female	10
	Female	10
	Male	10
	Female	10
	Female	10

```

289 -- 9. Quantity Analysis:
290 -- 9.a) Calculate the total quantity (quantity) sold
291 -- for each product line (product_line).
292 • select `product line`,sum(quantity) TotalCount
293 from walmart
294 group by `product line`
295 order by TotalCount desc;

```

Result Grid		
	product line	TotalCount
▶	Electronic accessories	971
	Food and beverages	952
	Sports and travel	920
	Home and lifestyle	911
	Fashion accessories	902

```

297 -- 10. VAT Analysis:
298 -- 10.a)Calculate the total VAT (VAT) collected for each city (city).
299 • select city, round(sum(`tax 5%`),2) VAT
300 from walmart
301 group by city
302 order by VAT desc;

```




Result Grid   Filter Rows: Export:  Wrap Cell Content: 

	city	VAT
▶	Naypyitaw	5265.18
	Yangon	5057.16
	Mandalay	5057.03

```

304 -- 10.b)Determine the city (city) with the highest total VAT collected.
305 • select city,`tax 5%` VAT
306 from walmart
307 order by VAT desc limit 1;

```

Result Grid   Filter Rows: Export:  Wrap Cell Content: 

	city	VAT
▶	Naypyitaw	49.65