

Task 3: SQL for Data Analysis

Dataset: [Bike Store Relational Database](#)

Tool: MySQL

Name: Sasindu Chanaka Piyumal.

Email ID: sasinduha@gmail.com

Introduction

The Bike Store dataset is a comprehensive relational database designed to simulate the operations of a bicycle retail business. It encompasses various key aspects of a real-world bike store, including product inventory, customer details, sales transactions, store locations, and staff information. This dataset is ideal for practicing and demonstrating SQL querying skills, data analysis, and business intelligence reporting.

The data is organized into multiple related tables, including:

- Customers – Contains information such as customer names, contact details, and location.

	customer_id	first_name	last_name	phone	email	street	city	state	zip_code
▶	1	Debra	Burks	NULL	debra.burks@yahoo.com	9273 Thorne Ave.	Orchard Park	NY	14127
	2	Kasha	Todd	NULL	kasha.todd@yahoo.com	910 Vine Street	Campbell	CA	95008
	3	Tameka	Fisher	NULL	tameka.fisher@aol.com	769C Honey Creek St.	Redondo Beach	CA	90278
	4	Daryl	Spence	NULL	daryl.spence@aol.com	988 Pearl Lane	Uniondale	NY	11553
	5	Charolette	Rice	(916) 381-6003	charolette.rice@msn.com	107 River Dr.	Sacramento	CA	95820
	6	Lyndsey	Bean	NULL	lyndsey.bean@hotmail.com	769 West Road	Fairport	NY	14450
	7	Latasha	Hays	(716) 986-3359	latasha.hays@hotmail.com	7014 Manor Station Rd.	Buffalo	NY	14215
	8	Jacqueline	Duncan	NULL	jacqueline.duncan@yahoo.com	15 Brown St.	Jackson Heights	NY	11372
	9	Genoveva	Baldwin	NULL	genoveva.baldwin@msn.com	8550 Spruce Drive	Port Washington	NY	11050
	10	Pamela	Newman	NULL	pamela.newman@gmail.com	476 Chestnut Ave.	Monroe	NY	10950
	11	Deshawn	Mendoza	NULL	deshawn.mendoza@yahoo.com	8790 Cobblestone Street	Monsey	NY	10952
	12	Robby	Sykes	(516) 583-7761	robby.sykes@hotmail.com	486 Rock Maple Street	Hempstead	NY	11550
	13	Lashawn	Ortiz	NULL	lashawn.ortiz@msn.com	27 Washington Rd.	Longview	TX	75604
	14	Garry	Espinoza	NULL	garry.espinoza@hotmail.com	7858 Rockaway Court	Forney	TX	75126
	15	Linnie	Branch	NULL	linnie.branch@gmail.com	314 South Columbia Ave.	Plattsburgh	NY	12901
	16	Emmitt	Sanchez	(212) 945-8823	emmitt.sanchez@hotmail.com	461 Squaw Creek Road	New York	NY	10002
	17	Caren	Stephens	NULL	caren.stephens@msn.com	914 Brook St.	Scarsdale	NY	10583
	18	Georgetta	Hardin	NULL	georgetta.hardin@aol.com	474 Chapel Dr.	Canandaigua	NY	14424
	19	Lizzette	Stein	NULL	lizzette.stein@yahoo.com	19 Green Hill Lane	Orchard Park	NY	14127
	20	Aleta	Shepard	NULL	aleta.shepard@aol.com	684 Howard St.	Sugar Land	TX	77478
	21	Tobias	Little	NULL	tobias.little@gmail.com	10 Silver Spear Dr.	Victoria	TX	77904

- Categories– Metadata about categories

	category_id	category_name
▶	1	Children Bicycles
	2	Comfort Bicycles
	3	Cruisers Bicycles
	4	Cyclocross Bicycles
	5	Electric Bikes
	6	Mountain Bikes
	7	Road Bikes

- Staff – Includes details of employees like salespersons and managers.

	staff_id	first_name	last_name	email	phone	active	store_id	manager_id
▶	1	Fabiola	Jackson	fabiola.jackson@bikes.shop	(831) 555-5554	1	1	NULL
	2	Mireya	Copeland	mireya.copeland@bikes.shop	(831) 555-5555	1	1	1
	3	Genna	Serrano	genna.serrano@bikes.shop	(831) 555-5556	1	1	2
	4	Virgie	Wiggins	virgie.wiggins@bikes.shop	(831) 555-5557	1	1	2
	5	Jannette	David	jannette.david@bikes.shop	(516) 379-4444	1	2	1
	6	Marcelene	Boyer	marcelene.boyer@bikes.shop	(516) 379-4445	1	2	5
	7	Venita	Daniel	venita.daniel@bikes.shop	(516) 379-4446	1	2	5
	8	Kali	Vargas	kali.vargas@bikes.shop	(972) 530-5555	1	3	1
	9	Layla	Terrell	layla.terrell@bikes.shop	(972) 530-5556	1	3	7
	10	Bernardine	Houston	bernardine.houston@bikes.shop	(972) 530-5557	1	3	7

- Products – Holds data about bicycles including brand, category, model, and price

	product_id	product_name	brand_id	category_id	model_year	list_price
▶	1	Trek 820 - 2016	9	6	2016	379.99
	2	Ritchey Timberwolf Frameset - 2016	5	6	2016	749.99
	3	Surly Wednesday Frameset - 2016	8	6	2016	999.99
	4	Trek Fuel EX 8 29 - 2016	9	6	2016	2899.99
	5	Heller Shagamaw Frame - 2016	3	6	2016	1320.99
	6	Surly Ice Cream Truck Frameset - 2016	8	6	2016	469.99
	7	Trek Slash 8 27.5 - 2016	9	6	2016	3999.99
	8	Trek Remedy 29 Carbon Frameset - 2016	9	6	2016	1799.99
	9	Trek Conduit+ - 2016	9	5	2016	2999.99
	10	Surly Straggler - 2016	8	4	2016	1549
	11	Surly Straggler 650b - 2016	8	4	2016	1680.99
	12	Electra Townie Original 21D - 2016	1	3	2016	549.99
	13	Electra Cruiser 1 (24-Inch) - 2016	1	3	2016	269.99
	14	Electra Girl's Hawaii 1 (16-inch) - 2015/2...	1	3	2016	269.99
	15	Electra Moto 1 - 2016	1	3	2016	529.99
	16	Electra Townie Original 7D EQ - 2016	1	3	2016	599.99
	17	Pure Cycles Vine 8-Speed - 2016	4	3	2016	429
	18	Pure Cycles Western 3-Speed - Women'...	4	3	2016	449
	19	Pure Cycles William 3-Speed - 2016	4	3	2016	449
	20	Electra Townie Original 7D EQ - Women's...	1	3	2016	599.99
	21	Electra Cruiser 1 (24-Inch) - 2016	1	1	2016	269.99
	22	Electra Girl's Hawaii 1 (16-inch) - 2015/2...	1	1	2016	269.99

- Brands – Provide metadata for product classification.

	brand_id	brand_name
▶	1	Electra
	2	Haro
	3	Heller
	4	Pure Cycles
	5	Ritchey
	6	Strider
	7	Sun Bicycles
	8	Surly
	9	Trek

- Orders – Records of customer purchases including dates, order status, and store/salesperson information.

	order_id	customer_id	order_status	order_date	required_date	shipped_date	store_id	staff_id
▶	1	259	4	2016-01-01	2016-01-03	2016-01-03	1	2
	2	1212	4	2016-01-01	2016-01-04	2016-01-03	2	6
	3	523	4	2016-01-02	2016-01-05	2016-01-03	2	7
	4	175	4	2016-01-03	2016-01-04	2016-01-05	1	3
	5	1324	4	2016-01-03	2016-01-06	2016-01-06	2	6
	6	94	4	2016-01-04	2016-01-07	2016-01-05	2	6
	7	324	4	2016-01-04	2016-01-07	2016-01-05	2	6
	8	1204	4	2016-01-04	2016-01-05	2016-01-05	2	7
	9	60	4	2016-01-05	2016-01-08	2016-01-08	1	2
	10	442	4	2016-01-05	2016-01-06	2016-01-06	2	6
	11	1326	4	2016-01-05	2016-01-08	2016-01-07	2	7
	12	91	4	2016-01-06	2016-01-08	2016-01-09	1	2
	13	873	4	2016-01-08	2016-01-11	2016-01-11	2	6
	14	258	4	2016-01-09	2016-01-11	2016-01-12	1	3
	15	450	4	2016-01-09	2016-01-10	2016-01-12	2	7
	16	552	4	2016-01-12	2016-01-15	2016-01-15	1	3
	17	1175	4	2016-01-12	2016-01-14	2016-01-14	1	3
	18	541	4	2016-01-14	2016-01-17	2016-01-15	1	3
	19	696	4	2016-01-14	2016-01-17	2016-01-16	1	2
	20	923	4	2016-01-14	2016-01-16	2016-01-17	1	2
	21	1250	4	2016-01-15	2016-01-16	2016-01-18	2	6
	22	1035	4	2016-01-16	2016-01-18	2016-01-17	1	2

- Order_Items – Line items in each order, specifying products, quantities, and prices.

	order_id	item_id	product_id	quantity	list_price	discount
▶	1	1	20	1	599.99	0.2
	1	2	8	2	1799.99	0.07
	1	3	10	2	1549	0.05
	1	4	16	2	599.99	0.05
	1	5	4	1	2899.99	0.2
	2	1	20	1	599.99	0.07
	2	2	16	2	599.99	0.05
	3	1	3	1	999.99	0.05
	3	2	20	1	599.99	0.05
	4	1	2	2	749.99	0.1
	5	1	10	2	1549	0.05
	5	2	17	1	429	0.07
	5	3	26	1	599.99	0.07
	6	1	18	1	449	0.07
	6	2	12	2	549.99	0.05
	6	3	20	1	599.99	0.1
	6	4	3	2	999.99	0.07
	6	5	9	2	2999.99	0.07
	7	1	15	1	529.99	0.07
	7	2	3	1	999.99	0.1
	7	3	17	2	429	0.1

- Stock – Information on product availability in each store.

	store_id	product_id	quantity
▶	1	1	27
	1	2	5
	1	3	6
	1	4	23
	1	5	22
	1	6	0
	1	7	8
	1	8	0
	1	9	11
	1	10	15
	1	11	8
	1	12	16
	1	13	13
	1	14	8
	1	15	3
	1	16	4
	1	17	2
	1	18	16
	1	19	4
	1	20	26
	1	21	24
	1	22	29

This dataset allows users to explore important business questions such as revenue trends, inventory management, customer behavior, and employee performance. It's widely used for learning SQL operations like JOIN, GROUP BY, AGGREGATE FUNCTIONS, and subqueries, making it a powerful resource for students and professionals in data-related fields.

Methodology

The following methodology was followed to analyze the Bike Store dataset and extract valuable business insights using MySQL:

1. Data Acquisition
The Bike Store dataset was downloaded from Kaggle, which provides a realistic simulation of a retail bicycle business, including information on customers, stores, staff, products, orders, and inventory.
2. Database Setup
A new database was created in MySQL to store and manage the dataset. This setup allowed efficient querying and relational data management.
3. Data Import
The individual dataset files (typically in .csv format) were imported into the MySQL database. Each file was imported as a separate table representing different entities such as customers, products, orders, order_items, staff, stores, brands, and categories

4. Data Exploration

Initial exploratory queries were run to understand the structure, relationships, and contents of each table. This included checking column names, data types, null values, foreign key connections, and sample data.

5. Scenario Analysis and Question Formulation

Based on business scenarios and use cases, a set of 25 meaningful and relevant analytical questions were designed. These questions cover various aspects such as sales trends, customer behavior, product performance, and staff efficiency.

6. Query Execution

The questions were answered by writing and executing SQL queries using MySQL. Queries made use of SQL concepts such as SELECT, JOIN, GROUP BY, ORDER BY, AGGREGATE FUNCTIONS, subqueries, and date functions.

7. Result Interpretation

The output of each query was analyzed to derive insights and support decision-making. Results were documented and, where applicable, visualized using basic tabular formats to enhance clarity.

Results

1. What are the total sales made by the store?

	total_sale
▶	8578988.880000794

2. What is the total number of orders placed?

	total_orders
▶	1615

3. Who are the top 5 customers by total purchase amount?

	first_name	last_name	amount
▶	Pamelia	Newman	37801.840000000004
	Abby	Gamble	37500.89
	Sharyn	Hopkins	37138.86
	Lyndsey	Bean	35857.859999999999
	Emmitt	Sanchez	34503.82

4. How many customers are there in each state?

	state	Num_Of_customers
▶	NY	1019
	CA	284
	TX	142

5. What is the total quantity of bikes sold by each store?

	store_id	count
▶	1	4722

6. Which customers have never placed an order?

	first_name	last_name	email
--	------------	-----------	-------

7. What is the average order value?

	avg(quantity)
▶	1.4989

8. What are the names and prices of all the bikes in stock?

	product_name	list_price
▶	Trek 820 - 2016	379.99
	Trek 820 - 2016	379.99
	Trek 820 - 2016	379.99
	Ritchey Timberwolf Frameset - 2016	749.99
	Ritchey Timberwolf Frameset - 2016	749.99
	Ritchey Timberwolf Frameset - 2016	749.99
	Surly Wednesday Frameset - 2016	999.99
	Surly Wednesday Frameset - 2016	999.99
	Surly Wednesday Frameset - 2016	999.99
	Trek Fuel EX 8 29 - 2016	2899.99
	Trek Fuel EX 8 29 - 2016	2899.99
	Trek Fuel EX 8 29 - 2016	2899.99
	Heller Shagamaw Frame - 2016	1320.99

Result 48 x

9. What are the top 10 most expensive bikes?

	product_name	list_price
▶	Trek Domane SLR 9 Disc - 2018	11999.99
	Trek Domane SLR 8 Disc - 2018	7499.99
	Trek Emonda SLR 8 - 2018	6499.99
	Trek Domane SL Frameset - 2018	6499.99
	Trek Domane SL Frameset Women's - 2018	6499.99
	Trek Silque SLR 8 Women's - 2017	6499.99
	Trek Silque SLR 7 Women's - 2017	5999.99
	Trek Domane SLR 6 Disc - 2017	5499.99
	Trek Domane SLR 6 Disc - 2018	5499.99
	Trek Domane SL 8 Disc - 2018	5499.99

10. Which product category has the highest number of items in stock?

	category_name	Product_Count
▶	Cruisers Bicycles	78

11. What is the total number of orders by each salesperson?

	first_name	last_name	Order_count
▶	Marcelene	Boyer	553
	Venita	Daniel	540
	Genna	Serrano	184
	Mireya	Copeland	164
	Kali	Vargas	88
	Layla	Terrell	86

12. How many employees are there in each store?

	store_name	Staff_count
▶	Santa Cruz Bikes	4
	Baldwin Bikes	3
	Rowlett Bikes	3

SQL: https://github.com/SasinduChanakaPiyumal/DATA-ANALYST-INTERNSHIP-/blob/main/DAY%203/bike_store.sql