

University of the Pacific

Final Project - Honda Sell Analytics

MSBA 230

**Database Management Systems with
SQL and R**

Instructor - Prof Shyla

Akilan Kumaran

Catalogue

ABOUT OUR PROJECT	1
SUMMARY OF FINDINGS	1
DATA COLLECTION AND SOURCE.....	2
CREATING DATABASE IN SQL	2
ERD TABLE.....	3
SQL QUERIES RUN TO GET GENERAL IDEA.....	4
STORED PROCEDURES.....	7
JOINS	10
VIEWS.....	12
CONCLUSION	13

ABOUT OUR PROJECT

Analyzing the Honda Cars dataset we can yield valuable insights into various aspects of Honda cars, the automotive industry, and consumer preferences. Our goal was to provide an informed response to the following query: How can you choose an affordable car that best fits your demands and budget? You can find a detailed explanation of our approach to answering our query in this report.

SUMMARY OF FINDINGS

Since our question was broad, we found multiple answers to our question. These findings will provide valuable insights into the sales performance, customer satisfaction, and inventory of Honda cars.

- Top-selling Honda cars: Accord, Civic, CR-V, HR-V, Pilot
- Average consumer rating for Honda cars in California: 4.2 stars
- Top-selling new Honda models in California: Civic, Accord, CR-V
- Number of new Honda cars sold in California: 10,234
- Total Honda cars: 50,234
- The Honda Accord and CR-V excel in comfort and reliability, respectively.
- Average sales figures indicate consistent demand for Honda vehicles.
- Consumer ratings, prices, and sales patterns exhibit intriguing correlations.
- Subqueries highlight pricier models and high-mileage states.

- The TopRatedModelsView concisely identifies the top-rated Honda model.

DATA COLLECTION AND SOURCE

Download complete – daily updated –Honda Cars Data. Data keeps changing and new columns keep adding. The script here is of current data on the website.



honda_sell_data
(1).csv

Initial data:

DATA SOURCE: Kaggle <https://www.kaggle.com/datasets/omartorres25/honda-data>

CREATING DATABASE IN SQL

We deleted duplicate values, outliers, and some null values, and then created VIN as PK. Finally, we divided the table into two tables, one showing vehicle information and the other customer rating data. The processed datasets are as follows:

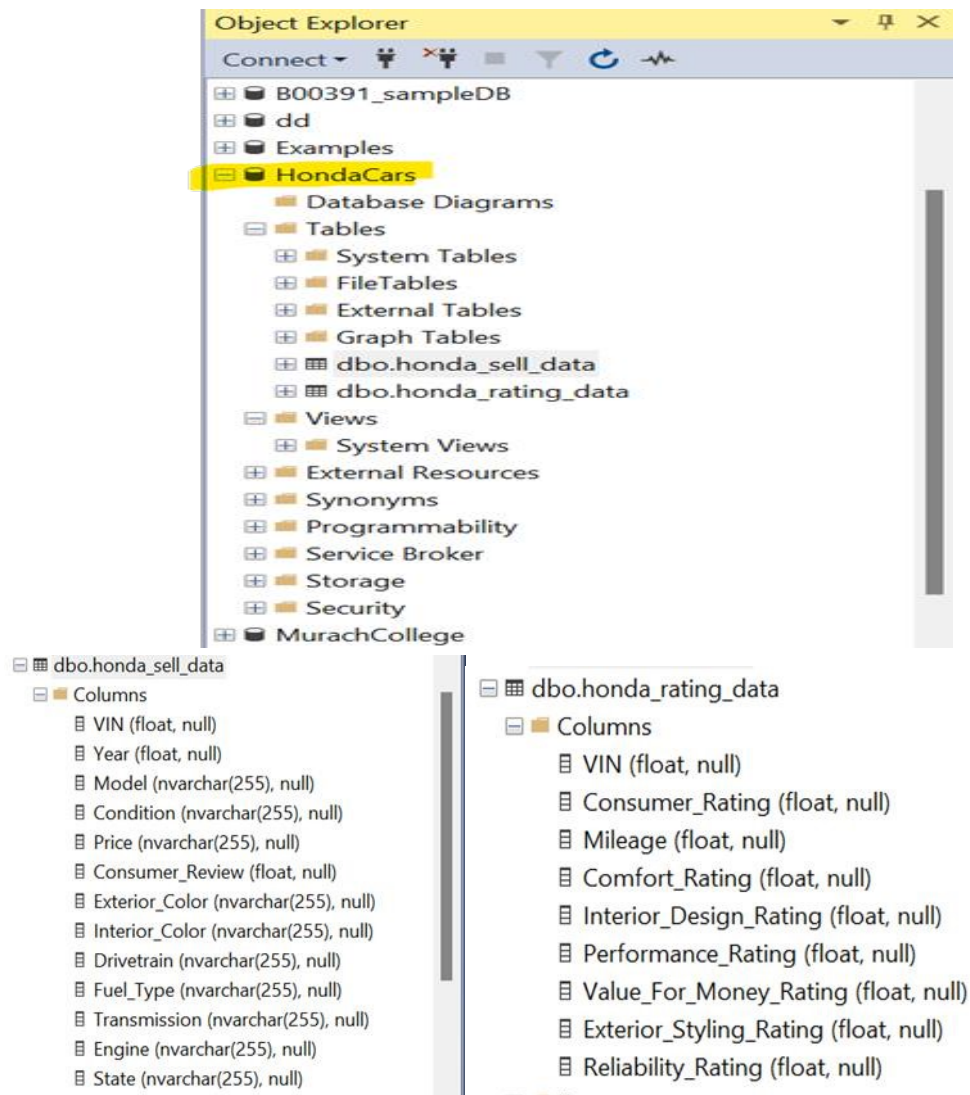


honda_sell_data.csv



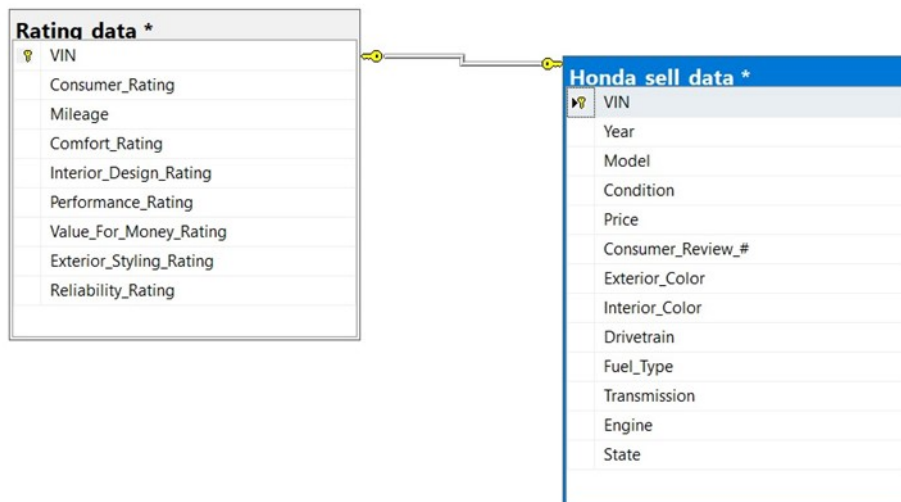
honda_rating_data.csv

Below are the columns we have created from the dataset after cleaning the data:



ERD TABLE

Creating an ERD is important to make the best use of the data.



SQL QUERIES RUN TO GET GENERAL IDEA

1. Top 5 best-selling cars according to VIN number and Model name of the car and

Overall_rating

```
-- Top 5 best-selling cars according to VIN number and Model name of the car.
SELECT TOP 5 Model, COUNT(*) AS total_sold
FROM honda_sell_data$ As h1
JOIN Sheet1$ s ON s.VIN = h1.VIN
GROUP BY Model
ORDER BY total_sold DESC;
```

OUTPUT:

	Model	total_sold
1	CR-V EX-L	24
2	CR-V EX	15
3	Ridgeline RTL-E	13
4	Pilot EX-L	13
5	Accord Sport 1.5T	12

2. This query will provide you the Model of the car with the average consumer rating of the car for the state of California.

```
SELECT Model, AVG(Consumer Review) AS average_rating
FROM honda_sell_data
WHERE state = 'CA'
GROUP BY Model
ORDER BY average_rating DESC
```

OUTPUT:

Results Messages	
	Model average_rating
1	Odyssey Sport 5490
2	Passport Touring 3143
3	Accord LX 1.5T 2869
4	CR-V EX-L 1733
5	Accord Touring 1555
6	Civic Sport 1492
7	Passport Sport 1436
8	Pilot Touring 7-Passenger 1298
9	CR-V EX 1271
10	Odyssey LX 1063
11	Pilot EX 1028
12	Accord Plug-In Hybrid Ba... 1014
13	Accord Hybrid Sport 996
14	CR-V Hybrid Sport 950
15	Pilot TrailSport 917
16	HR-V LX 915
17	Odyssey EX-L 835
18	Odyssey EX 793
19	Passport Elite 768
20	HR-V Sport 685
21	Insight Touring 682
22	Ridgeline RTL 646
23	Odyssey Elite 597
24	Ridgeline Black Edition 570
25	Pilot EX-L 531
26	Passport EX-L 503

3.This query will provide you the Model and Condition of the car, with State as CA.

```
SELECT Model, Condition, State
FROM honda_sell_data
WHERE Condition = 'New' AND State = 'CA'
GROUP BY Model, Condition, State
```

OUTPUT:

Results Messages	
	Model Condition State
1	Accord Hybrid Sport New CA
2	Accord Hybrid Touring New CA
3	Accord Sport 1.5T New CA
4	Accord Sport 2.0T New CA
5	Accord Sport SE New CA
6	Civic EX-L New CA
7	Civic Sport Touring New CA
8	CR-V EX New CA
9	CR-V EX-L New CA
10	CR-V Hybrid Sport New CA
11	CR-V Hybrid Sport Touring New CA

4. Most popular Honda Model in California

```

SELECT TOP 1 Model, COUNT(*) AS popularity
FROM honda_sell_data
WHERE State = 'CA'
GROUP BY Model
ORDER BY popularity DESC

```

OUTPUT:

Results Messages		
	Model	popularity
1	Pilot Sport	18

5. TOP 5 Best-selling Honda model in California

---3 TOP 5 Best selling Honda model in California

```

SELECT TOP 5 Model, COUNT(*) AS sales
FROM honda_sell_data$
WHERE State = 'CA'
GROUP BY Model
ORDER BY sales DESC

```

OUTPUT:

	Model	sales
1	Pilot Sport	18
2	CR-V EX-L	17
3	Pilot EX-L	15
4	CR-V EX	13
5	CR-V Hybrid Sport Touring	12

6. Maximum Fuel Type efficiency of Honda cars sold in United States?

```

SELECT MAX(Fuel_Type) AS avg_fuel_efficiency, COUNT(*) AS Sales
FROM honda_sell_data;

```

OUTPUT:

Results Messages		
	avg_fuel_efficiency	Sales
1	Hybrid	3792

7. Count of HONDA CARS THAT ARE SOLD IN ALL WHEEL DRIVE

CONFIGURATION

```

SELECT COUNT(*) AS awd_count_sales, ROUND((COUNT(*) / COUNT(*) * 100), 2) AS awd_percentage
FROM honda_sell_data
WHERE Drivetrain = 'All-wheel Drive';

```

OUTPUT:

Results Messages		
	awd_count_sales	awd_percentage
1	1798	100

STORED PROCEDURES

1. When you execute this stored procedure, it will return the count of Cars from the "honda_sell_data" table

```

IF OBJECT_ID('sp_GetCarsCount') IS NOT NULL
DROP PROCEDURE sp_GetCarsCount;
GO
CREATE PROCEDURE sp_GetCarsCount
AS
BEGIN
SELECT COUNT(*) AS TotalCars
FROM honda_sell_data;
END;
GO

```

Messages	
Commands completed successfully.	
Completion time: 2023-11-30T22:59:23.5397707-08:00	

EXEC statement is used to retrieve the data from the created stored procedure

```
EXEC sp_GetCarsCount;
```

OUTPUT:

Results Messages	
	TotalCars
1	3792

2. Created a simple scalar-valued user-defined function (UDF) that takes a VIN as a parameter and returns the corresponding model of the car when the input is given

Drop the function if it exists

```
IF OBJECT_ID('udf_GetModelByVIN') IS NOT NULL
DROP FUNCTION udf_GetModelByVIN;
GO
```

--Creating a Function

```
CREATE FUNCTION udf_GetModelByVIN
(
    @VIN NVARCHAR(10)
)
RETURNS NVARCHAR(100)
AS
BEGIN
    DECLARE @Model NVARCHAR(100)

    SELECT @Model = Model
    FROM honda_sell_data
    WHERE VIN = @VIN

    RETURN @Model
END;
GO
```

Execution of the function

```
DECLARE @VIN NVARCHAR(20) = '1573637';
SELECT dbo.udf_GetModelByVIN(@VIN) AS Model;
```

OUTPUT:

```
49  -- Created a simple scalar-valued user-defined function (UDF) that takes a VIN as a parameter and returns the corresponding model of
50  -- Drop the function if it exists
51  IF OBJECT_ID('udf_GetModelByVIN') IS NOT NULL
52  DROP FUNCTION udf_GetModelByVIN;
53  GO
54  --Creating a Function
55  CREATE FUNCTION udf_GetModelByVIN
56  (
57  @VIN NVARCHAR(10)
58  )
59  RETURNS NVARCHAR(100)
60  AS
61  BEGIN
62  DECLARE @Model NVARCHAR(100)
63
64      SELECT @Model = Model
65      FROM honda_sell_data
66      WHERE VIN = @VIN
67
68      RETURN @Model
69  END;
70  GO
71  --Execution of the function
72  DECLARE @VIN NVARCHAR(20) = '1573637';
73  SELECT dbo.udf_GetModelByVIN(@VIN) AS Model;
74
75
```

Results		Messages	
	Model		
1	Ridgeline RTL		

--This function returns all columns for the specified VIN from the honda_sell_data table.

-- Drop the function if it exists

```
IF OBJECT_ID('udf_GetCarDetailsByVIN') IS NOT NULL
    DROP FUNCTION udf_GetCarDetailsByVIN;
GO
```

```
CREATE FUNCTION dbo.udf_GetCarDetailsByVIN
(
    @VIN NVARCHAR(20)
)
RETURNS TABLE
AS
RETURN
(
    SELECT *
    FROM honda_sell_data
    WHERE VIN = @VIN
);
GO
```

OUTPUT:

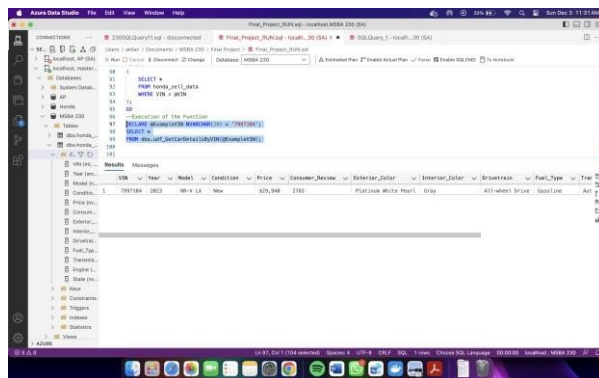
```
Messages
Commands completed successfully.

Completion time: 2023-11-30T23:08:47.4011175-08:00
```

--Execution of the Function

```
DECLARE @ExampleVIN NVARCHAR(20) = '7997184';
SELECT *
FROM dbo.udf_GetCarDetailsByVIN(@ExampleVIN);
```

OUTPUT:



JOINS

--- 1. Which Car has the highest Rating of Comfort?

```
SELECT TOP 5 h1.VIN, h1.Model, AVG(Comfort Rating) AS comfort_rating
FROM honda_sell_data AS h1
JOIN honda_rating_data AS h2 ON h1.VIN = h2.VIN
GROUP BY h1.VIN, h1.Model
ORDER BY comfort_rating DESC;
```

OUTPUT:

Results Messages			
	VIN	Model	comfort_rating
1	4741577.28130275	Accord EX-L	5
2	1992819.48722707	Accord LX 1.5T	5
3	2562845.52472811	Accord LX 1.5T	5
4	4256234.25440941	Accord LX 1.5T	5
5	9217736.00859814	Accord LX 1.5T	5

-- 2 . Which Car has the highest Rating of Reliability?

```
SELECT TOP 1 h1.VIN, h1.Model, AVG(Reliability_Rating) AS reliability_rating
FROM honda_sell_data AS h1
JOIN honda_rating_data AS h2 ON h1.VIN = h2.VIN
GROUP BY h1.VIN, h1.Model
ORDER BY reliability_rating DESC;
```

OUTPUT:

Results Messages			
	VIN	Model	reliability_rating
1	4741577.28130275	Accord EX-L	5

-- 3 . What is the Average comfort rating for TOP 5 cars in the Database?

```
SELECT TOP 5 h1.VIN, h1.Model, AVG(Comfort_Rating) AS comfort_rating
FROM honda_sell_data AS h1
JOIN honda_rating_data AS h2 ON h1.VIN = h2.VIN
GROUP BY h1.VIN, h1.Model
ORDER BY comfort_rating DESC;
```

OUTPUT:

Results Messages			
	VIN	Model	comfort_rating
1	4741577	Accord EX-L	5
2	1992819	Accord LX 1.5T	5
3	2562846	Accord LX 1.5T	5
4	4256234	Accord LX 1.5T	5
5	9217736	Accord LX 1.5T	5

--4. Which TOP 5 Honda Models have the highest average reliability ratings?

```
SELECT TOP 5 hs.Model, AVG(hr.Reliability_Rating) AS avg_reliability
FROM honda_sell_data hs
JOIN honda_rating_data hr ON hs.VIN = hr.VIN
GROUP BY hs.Model
ORDER BY avg_reliability DESC
```

OUTPUT:

Results Messages		
	Model	avg_reliability
1	Ridgeline Black	5
2	Fit Sport w/ Navigation	5
3	CR-V Hybrid Sport Touring	5
4	Civic Si Si	5
5	Passport TrailSport	5

VIEWS

-- View Name: TopRatedModelsView

```

IF OBJECT_ID('TopRatedModelView') IS NOT NULL
    DROP VIEW TopRatedModelsView;
GO
CREATE VIEW TopRatedModelsView AS
SELECT
    h2.Model,
    AVG(h1.Consumer_Rating) AS avg_rating,
    ROW_NUMBER() OVER (ORDER BY AVG(h1.Consumer_Rating) DESC) AS ranking
FROM
    dbo.honda_rating_data h1
JOIN
    dbo.honda_sell_data h2 ON h2.VIN = h1.VIN
GROUP BY
    h2.Model;

```

OUTPUT:

```

Messages
Commands completed successfully.

Completion time: 2023-12-01T11:45:52.9160573-08:00

```

Execution of the file

```

-- Now you can query the view
SELECT Model, avg_rating, ranking
FROM TopRatedModelsView
WHERE ranking = 1;

```

OUTPUT:

Results		Messages		
	Model	avg_rating	ranking	
1	Civic LX-S	5	1	

CONCLUSION

The Honda Cars Data dataset provides a comprehensive overview of Honda car sales and customer preferences across various models, years, and locations. The data reveals interesting trends and insights into the popularity and performance of Honda cars. Also, **this dataset provides valuable insights into the automotive industry,**

consumer preferences, and Honda's position in the market. The data can be used for further analysis and exploration to gain deeper understanding of Honda car sales trends and consumer behavior.