[Analyses Sheet](https://docs.google.com/spreadsheets/d/1596h7jRJw2CN2XYImiE0Sgcj3kSIWUR2O61BkhVEJ8o/edit#gid=1863667694) 1

[Analyses sheet 2](https://docs.google.com/spreadsheets/d/1rt1VBEThZI8GAgYistiWQ6ECT3tMJ0yUNWAqlSSKCn8/edit#gid=1513423556)

* [Sheet- Q10](https://docs.google.com/spreadsheets/d/1XGoN8L9r-m7KrzS299Oz_AV3Q72xJapdIhrQEjFFvtU/edit#gid=1705932408): Find the first IPL match in which Sachin was the man of the match [INDEX, MATCH]
  1. Index [doc](https://support.google.com/docs/answer/3098242?hl=en)
  2. Let E6 = SR Tendulkar
  3. F6 =INDEX('IPL Matches 2008-2020'!A2:R817,MATCH(E6,'IPL Matches 2008-2020'!D2:D817,0),1)
* If you want to see all the data for which Sachin was the man of the match
  1. =FILTER('IPL Matches 2008-2020'!A1:R817,'IPL Matches 2008-2020'!D1:D817=E6)
* HLOOKUP: just like VLOOKUP
  1. Given the table in sheet Q11

| Name | SR Tendulkar | V Sehwag | V Kohli |
| --- | --- | --- | --- |
| Matches | 200 | 97 | 98 |
| 100's | 20 | 22 | 21 |
| 50's | 32 | 35 | 25 |
| 6's | 100 | 25 | 54 |
| 4s | 125 | 20 | 14 |

* 1. Put player name in H7. For eg: H7 = V Kohli
  2. If you want to see number of 100s, H8 = HLOOKUP(H7,A6:D11,3,FALSE)
* Create a sheet with match-ID in rows, over-number in columns, sum of runs and extras for each combination. Can we filter by innings too? [Pivot Table]
  1. [1] Goto Ball-by-Ball sheet—> Insert—> Pivot Table —> Existing sheet —> Selection grid —> Q12!A4 —> Create
  2. [2] Rows—> Add—> ID Columns —> Add —> Over Values—> Add —> Total\_runs Values—> Add—> Extras FILTER—> Innings
  3. Same as LoD in Tableau; group-by in SQL
* Get number of overs, Runs, Extras and Economy and per bowler [Pivot, Calculated Field]
  1. [1] Goto Ball-by-Ball sheet—> Insert—> Pivot Table —> Existing sheet —> Selection grid —> Q13!A4 —> Create
  2. [2] Rows—> Add—> bowler Values—> Add —> total\_runs Values—> Add—> Extras Values—> Add-> Calculated Field —> Formula: =COUNTUNIQUE(id, over); Summarised by “Custom”
  3. [3] Goto the sheet change name of Calculated Field 1 to “Overs”
  4. Economy = cell[sum(total runs)]/cell[overs]
* Colour all rows where the city=Bangalore [Conditional Formatting]
  1. [1] In matches table —> Format—> Conditional Formatting —>
  2. [2] Apply to Range: A2:Q817 Formula —>
  3. Under format cells if -> Custom Formula —>
     1. =$B2="Bangalore"
  4. [3] Sort an fIlter by Color can be performed if need be. Another example: =MONTH($C2)=5
* SIMPLE CHARTS on Excel
  1. Use Samplestore [datasource](https://docs.google.com/spreadsheets/d/1sM1Kjgy99k2tHBtKP5OXcpQLz7SC5yGr/edit#gid=432276125)
* Q1:Scatter plot between discount vs profit and discount vs sale-price. Show trend lines

[1] Create a new sheet, Insert —> Chart

[2] Chart type: Scatter

[3] X-axis: Orders!T1:T9964 (Discount)

Series: Orders!U1:U9964 (Profit)

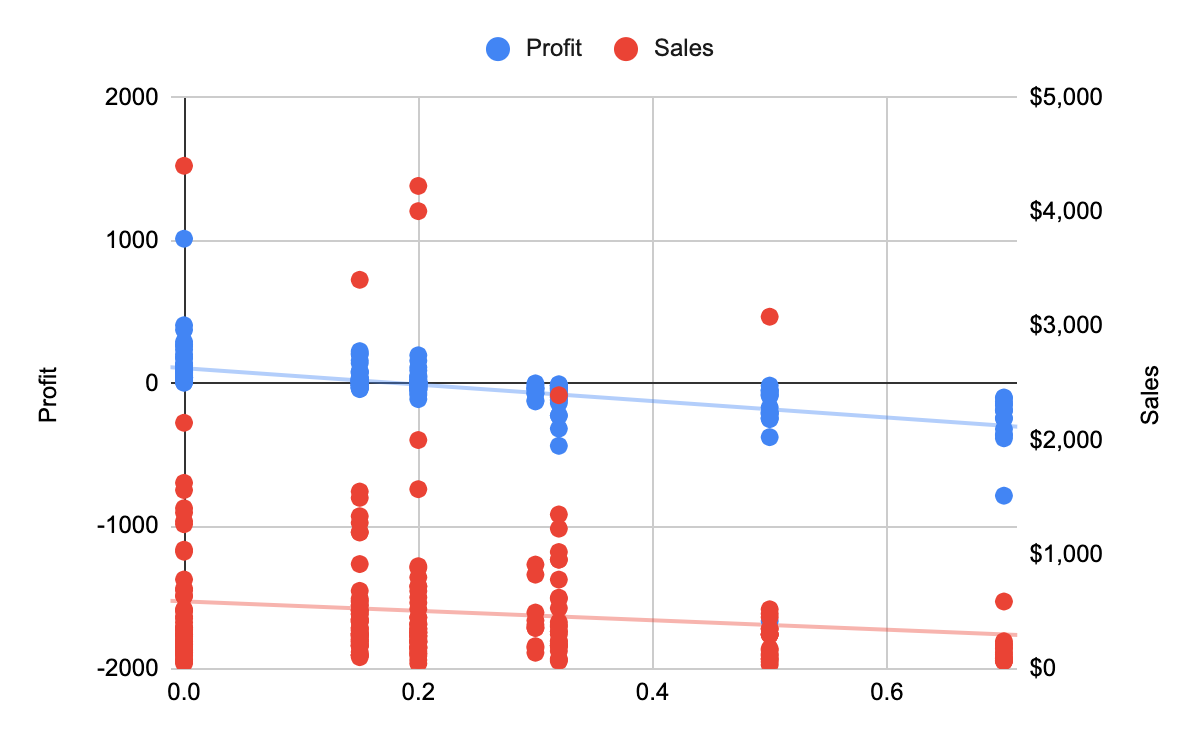
Series: Orders!R1:R9964. (Sales)

[Tick] User Row 1 as headers

[4] Chat Editor —> Customise —> Series—> Data Series 2—> Axis —> Right axis;

[5] Chat Editor —> Customise —> Series—> Apply to all series —> Trendline —> Type: Linear

[6] Customize—> Chart and axis titles—> Change title and vertical (dual) axis names



The problem here is that for each discount value on the x axis, there are multiple overlapping points for sales and profits. So instead of this can we just take one avg value for each discount point?

Q1a: Can you use a pivot table to plot this better?

* Create pivot table with the datasource
* Drag Discount to rows
* Drag profit to values -> make it average
* Drag sales to values -> make it average
* Now choose the pivot table -> insert -> chart

Q2. Histogram of sale prices

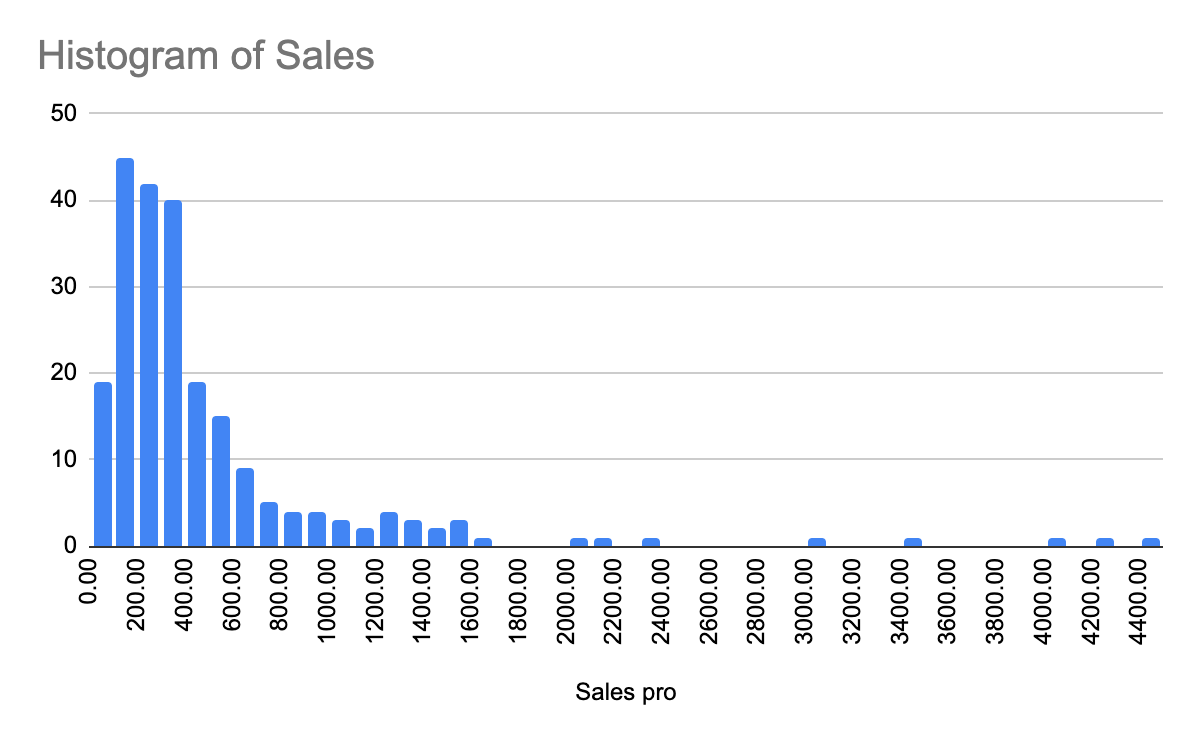
[1] Insert Charts—> Setup—> Type: Histogram -> Under X Axis, select range of sales R:R from orders

[1a] untick “Use column R as labels”

[2] Customize —> Chart & Axis Title —> Horizontal axis —> Sale price

Vertical axis —> Count

[3] Customize —> Histogram—>Bucketsize=50



Q3. Geo chart: sales and profits per state

[1] Pivot table from Orders:

Rows—> State

Values—>SUM(Sales); SUM(Profits)

[2] Ctrl+a on the data / click on a blank cell -> Insert Charts—> Setup—> Type: Map with markers

[2] Customize—>

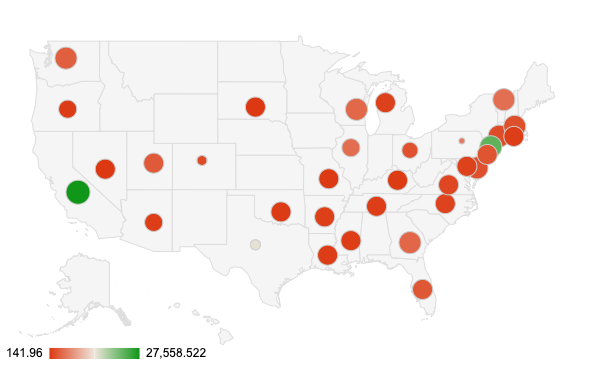
Geo—> Region=US

[3] Setup—>

Region: A1:A50

Color: B1:B50 [Sales]

Size: C1:C50 [Profit]



Q4. Candle chart: Show min, max, average and median sales per state

[1] Create pivot table from Orders

Rows—> State

Values—>MAX(sales), MIN(Sales, AVERAGE(Sales), MEDIAN(Sales)

[2] Insert Chart —> Type: candlestick

X-axis: A1:A50

Low:C1:C50 [min]

Open:D1:D50 [Average]

Close:E1:E50 [Median]

High: B1:B50 [max]

[3] Customize—> Vertical Axis —> Log-scale

Log 1 (base 10) = 0

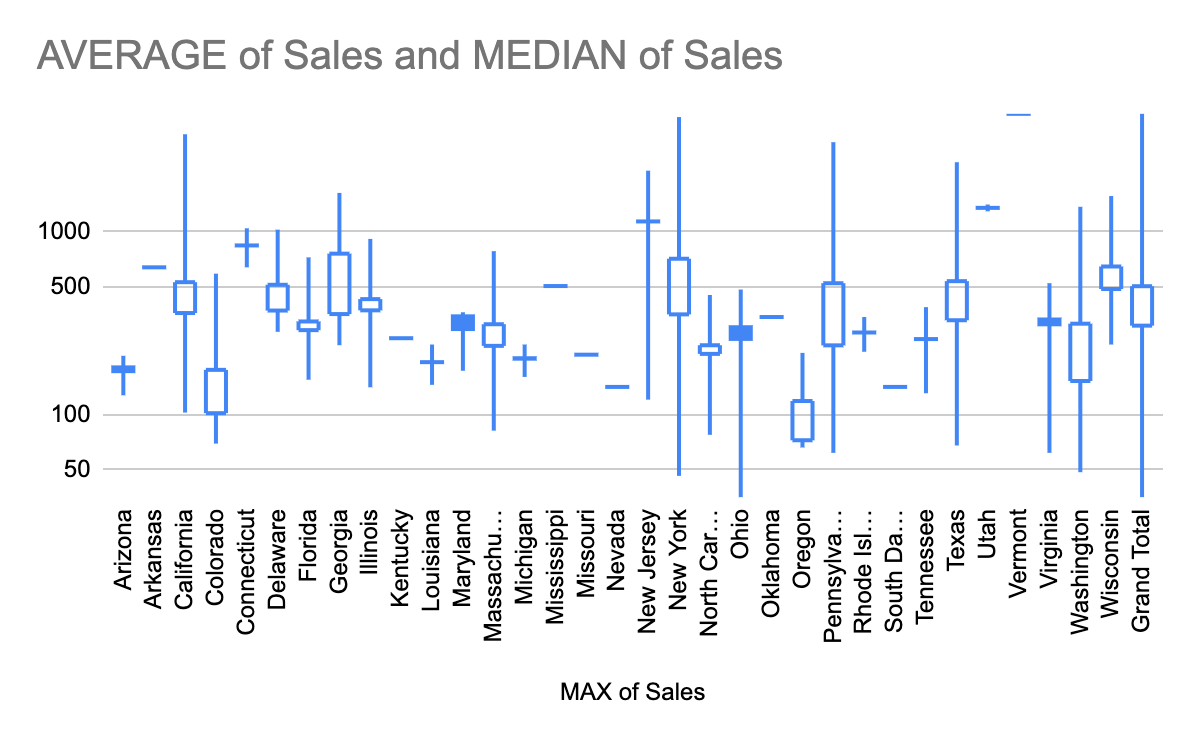
Log 10 (base 10) = 1

Log 100 (base 10) = 2

Log 1000 (base 10) = 3

The reason to use logarithmic scales is **to resolve an issue with visualizations that skew towards large values in a dataset**.

A logarithmic scale is a nonlinear scale often used when analyzing a large range of quantities. Instead of increasing in equal increments, **each interval is increased by a factor of the base of the logarithm**.



Q5: Format your table to read data easily using alternating colours for each row

[1] Select the table using Cntrl+A

[2] Format—> Alternating Colors—> Pick a color palette of your choice

Q6. Automate repetitive actions: Take the top 10 orders from California and create a new sheet with that data.

[1] Extensions —> Macros —> Record Macro —> <Complete your task> —> Save Macro with a name

**If you are using macro for the first time , they may ask you for authroize permission to change you sheet, give permission**

Task:

1. Filter order by state
2. Go to sales, sort sales by desc

[2] Test the macro by running it from the Orders sheet.

**If you go to extensions -> macros -> now you can see your new macro created**

**If the data you stored is in xlsx format in google sheets , you wont see extension ->,go to file -> save as google sheets**

Q7. Perform ML tasks using Google sheets: Fill missing data

[1] Open HousePrice.csv in a new sheet

<https://drive.google.com/file/d/1rusCRcZs-C3q4yCgxD0FWJseleNQ1CW_/view?usp=sharing>

Source:<https://www.kaggle.com/competitions/house-prices-advanced-regression-techniques/data>

[2] Create a copy of that sheet -> Randomly delete some SalePrice values at random. We will fill them using ML

[2] Extensions—> Addons—> Search for Simple ML —> Install —> give permissions

[3] Refresh the page —> Extensions —> Simple ML for sheets —>Start —> Predict missing values —> Choose SalePrice column—> Give it time.

[4] Source columns —> Select all columns but ID: Which can help us predict the sale price

[5] Choose the Learning algorithm -> gradient boosted trees

[6] how far is my predicted value with my actual value ? Predicted - actual/actual

Other performance metrics: MSE, RMSE, MAE