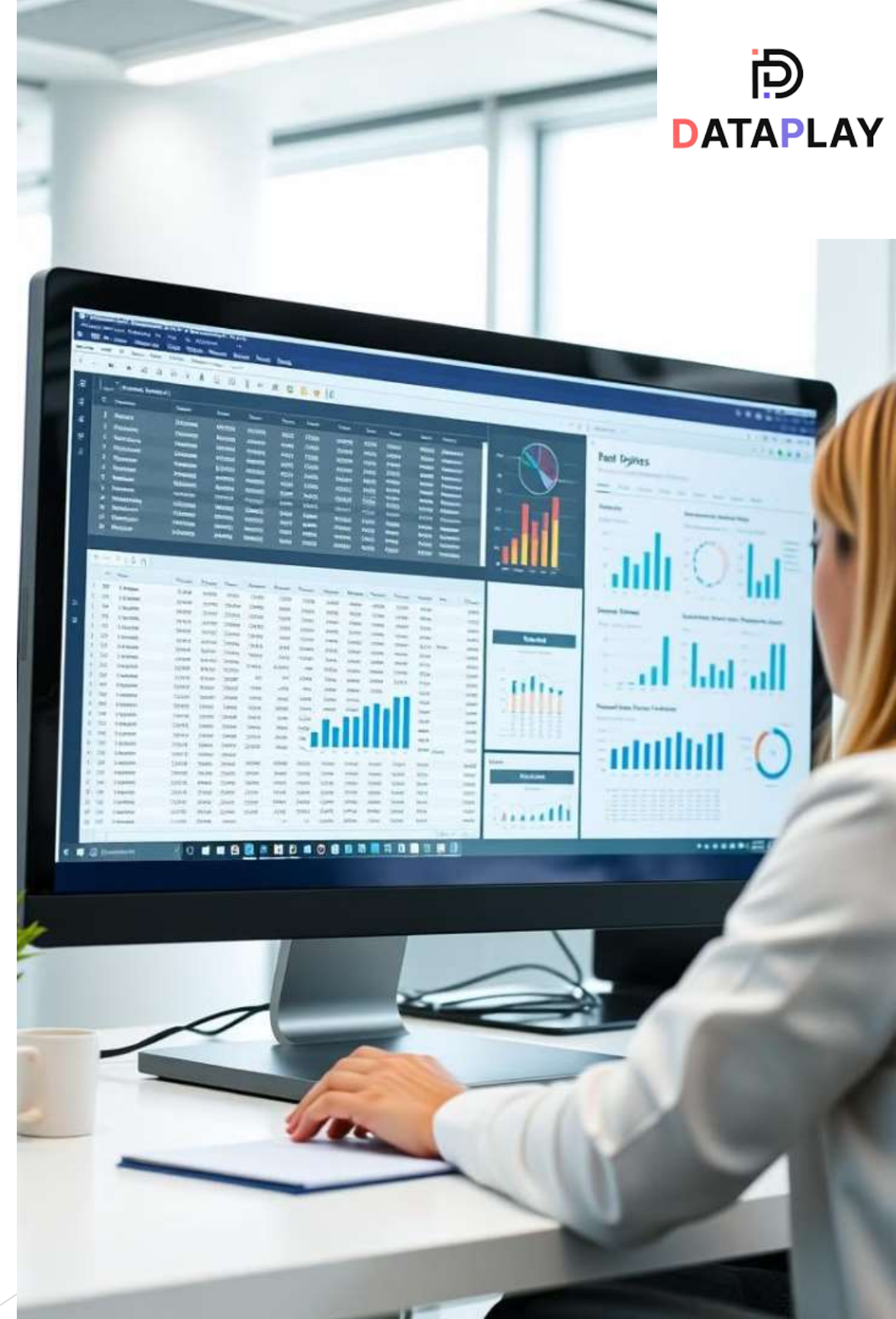


Gender Prediction Using Excel

This presentation explores how to predict gender using Excel, it offers a practical and accessible approach to data analysis.

BY: Kanupriya Nailwal



Problem Statement:

Unveil the gender mystery behind each name with the power of Excel.

Name	Gender
Meghamala	Female
Pragya	Female
Kranti	Female
Tulika	Female
Aarushi	Female
Pratigya	Female
Madhavi	Female
Charusila	Female
Manmayi	Female
Mahajabeen	Female
Nidhyathi	Female
Nainika	Female
Madirakshi	Female
Harijatha	Female
Rupal	Female
Eka	Female
Ilampirai	Female
Shrivalli	Female
Oma	Female
Kali	Female
Gudiya	Female
Neelabja	Female

Data Splitting: Training and Validation :

This dataset consists of 3001 rows of data. Split data into 70% for training and 30% for validation using stratified splitting technique.

Tools & Formula Used

1 IF()

To predict the gender.

If female% is greater than male%, predict female, otherwise male.

2 Right()

It returns the last character of the name .

3 Pivot Table

To summarize data and generate insights.

4 VLOOKUP()

To find the gender value based on the last letter from test data and use that data to predict the Gender based on the new names in the test data.

Methodology:

- Extract the **last letter** from each name from training data.
- Check how frequently the letter occurs, corresponding to the gender. For instance: The letters '**a**' & '**i**' most frequently occur when the person is **Female**.
- Use this information to predict the gender.

Name ▼	Gender ▼	LastName ▼
Meghama	Female	a
Pragya	Female	a
Kranti	Female	i
Tulika	Female	a
Aarushi	Female	i
Pratigya	Female	a
Madhavi	Female	i
Charusila	Female	a
Manmayi	Female	i
Mahajabe	Female	n
Nidhyathi	Female	i
Nainika	Female	a
Madirakst	Female	i
Harijatha	Female	a
Rupal	Female	l
Eka	Female	a
Ilampirai	Female	i
Shrivalli	Female	i
Oma	Female	a
Kali	Female	i
Gudiya	Female	a
Neelabja	Female	a
Sarbani	Female	i
Sharvani	Female	i
Tripuri	Female	i
Naisha	Female	a

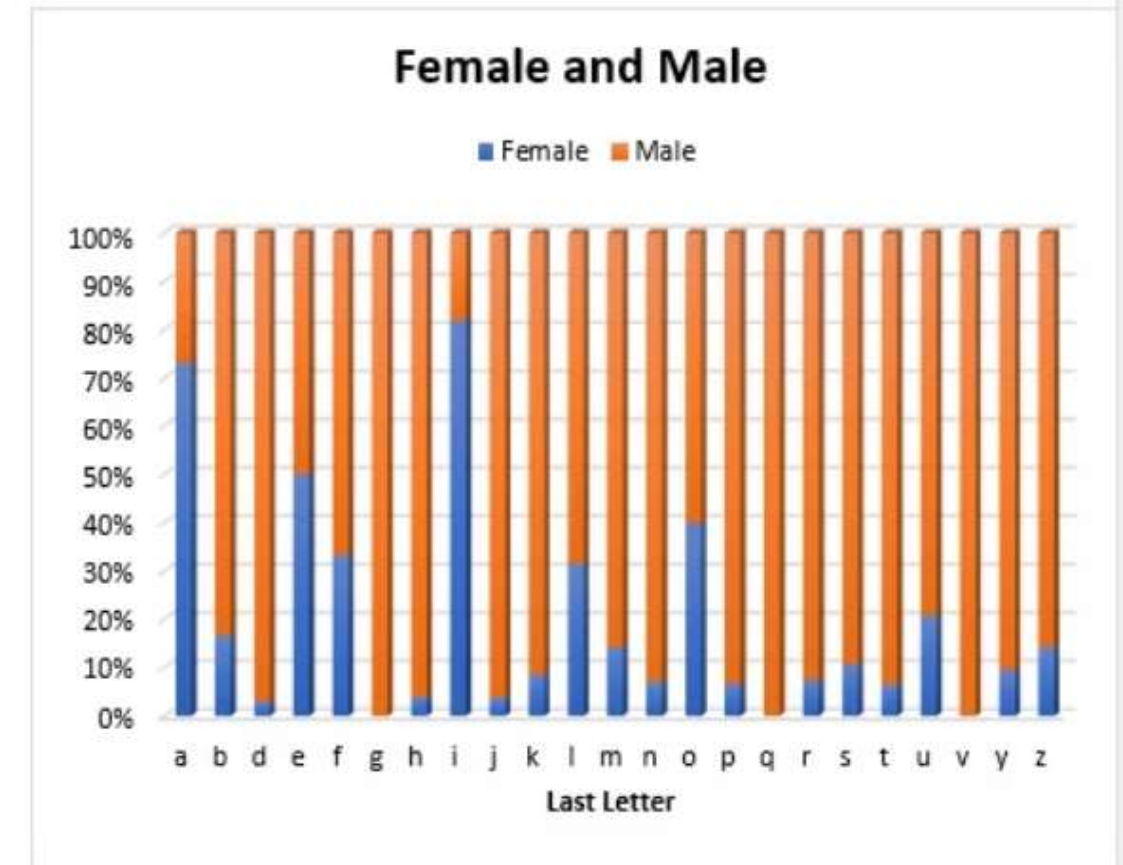
Pivot Table

► Pivot tables are powerful for analyzing data by gender, summarizing and visualizing trends. Here a pivot table is created based on training dataset.

Last Letter ▾	Female	Male
a	73.04%	26.96%
b	16.67%	83.33%
d	2.86%	97.14%
e	50.00%	50.00%
f	33.33%	66.67%
g	0.00%	100.00%
h	3.94%	96.06%
i	81.93%	18.07%
j	3.70%	96.30%
k	8.70%	91.30%
l	31.58%	68.42%
m	14.29%	85.71%
n	7.11%	92.89%
o	40.00%	60.00%
p	6.67%	93.33%
q	0.00%	100.00%
r	7.41%	92.59%
s	10.71%	89.29%
t	6.49%	93.51%
u	20.83%	79.17%
v	0.00%	100.00%
y	9.52%	90.48%

Training data analysis:

	A	B	C	D	E	F	G	H	I	J	K	L
1	Name	Gender	LastName									
2	Meghama	Female	a	Count of Nam Column Labe								
3	Pragya	Female	a	Last Letter	Female	Male						
4	Kranti	Female	i	a	73.04%	26.96%						
5	Tulika	Female	a	b	16.67%	83.33%						
6	Aarushi	Female	i	d	2.86%	97.14%						
7	Pratigya	Female	a	e	50.00%	50.00%						
8	Madhavi	Female	i	f	33.33%	66.67%						
9	Charusila	Female	a	g	0.00%	100.00%						
10	Manmayi	Female	i	h	3.94%	96.06%						
11	Mahajabe	Female	n	i	81.93%	18.07%						
12	Nidhyathi	Female	i	j	3.70%	96.30%						
13	Nainika	Female	a	k	8.70%	91.30%						
14	Madirakst	Female	i	l	31.58%	68.42%						
15	Harijatha	Female	a	m	14.29%	85.71%						
16	Rupal	Female	l	n	7.11%	92.89%						
17	Eka	Female	a	o	40.00%	60.00%						
18	Ilampirai	Female	i	p	6.67%	93.33%						
19	Shrivalli	Female	i	q	0.00%	100.00%						
20	Oma	Female	a	r	7.41%	92.59%						
21	Kali	Female	i	s	10.71%	89.29%						
22	Gudiya	Female	a	t	6.49%	93.51%						
23	Neelabja	Female	a	u	20.83%	79.17%						
24	Sarbani	Female	i	v	0.00%	100.00%						
25	Sharvani	Female	i	y	9.52%	90.48%						
26	Tripuri	Female	i	z	14.29%	85.71%						
27	Naisha	Female	a	Grand Total	46.34%	53.66%						



Prediction On Validation Data

Using the VLOOKUP function and using the pivot table as reference the gender was predicted and using IF function the probabilities were compared.

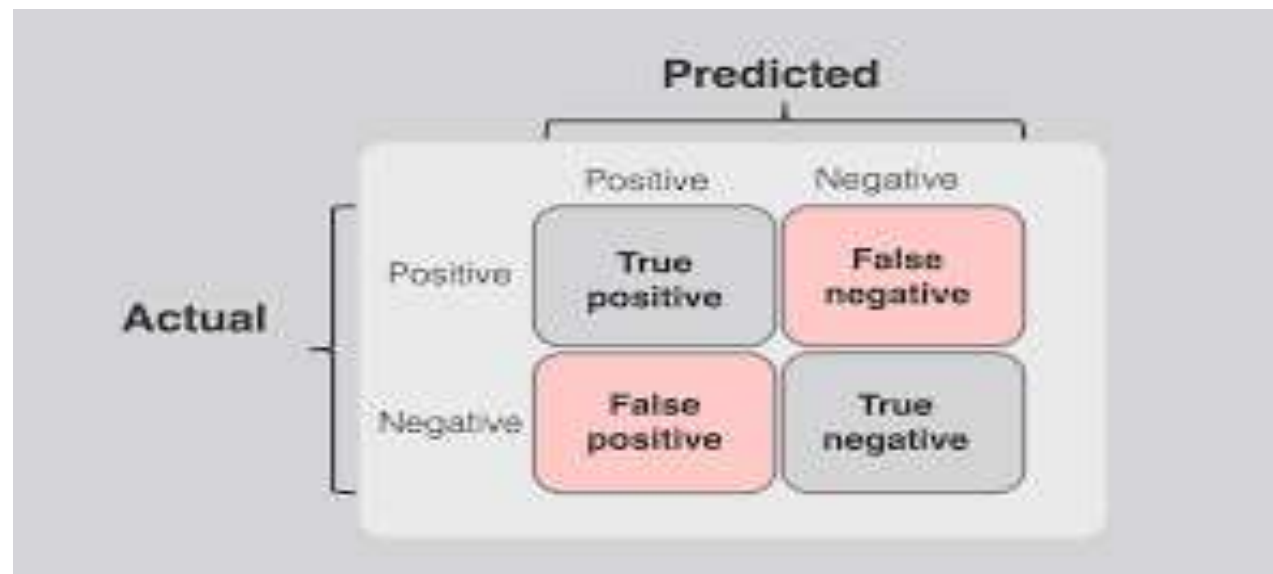
	A	B	C	D	E	F	G
1	Column1 ▾	Column2 ▾	Column3 ▾	Column4 ▾	Column5 ▾	Column6 ▾	Column7 ▾
2	Name	Gender	LastName	Gender_M	Gender_F	Gender_predicted	RuleBasedModel(Accuracy)
3	Meghamala	Female	a	0.269647696	0.7303523	Female	TRUE
4	Pragya	Female	a	0.269647696	0.7303523	Female	TRUE
5	Kranti	Female	i	0.180693069	0.81930693	Female	TRUE
6	Tulika	Female	a	0.269647696	0.7303523	Female	TRUE
7	Aarushi	Female	i	0.180693069	0.81930693	Female	TRUE
8	Pratigya	Female	a	0.269647696	0.7303523	Female	TRUE
9	Madhavi	Female	i	0.180693069	0.81930693	Female	TRUE
10	Charusila	Female	a	0.269647696	0.7303523	Female	TRUE
11	Manmayi	Female	i	0.180693069	0.81930693	Female	TRUE
12	Mahajabeen	Female	n	0.92893401	0.07106599	Male	FALSE
13	Nidhyathi	Female	i	0.180693069	0.81930693	Female	TRUE
14	Nainika	Female	a	0.269647696	0.7303523	Female	TRUE
15	Madirakshi	Female	i	0.180693069	0.81930693	Female	TRUE
16	Harijatha	Female	a	0.269647696	0.7303523	Female	TRUE
17	Rupal	Female	l	0.684210526	0.31578947	Male	FALSE
18	Eka	Female	a	0.269647696	0.7303523	Female	TRUE
19	Ilampirai	Female	i	0.180693069	0.81930693	Female	TRUE
20	Shrivalli	Female	i	0.180693069	0.81930693	Female	TRUE
21	Oma	Female	a	0.269647696	0.7303523	Female	TRUE
22	Kali	Female	i	0.180693069	0.81930693	Female	TRUE
23	Gudiya	Female	a	0.269647696	0.7303523	Female	TRUE
24	Neelabja	Female	a	0.269647696	0.7303523	Female	TRUE
25	Sarbani	Female	i	0.180693069	0.81930693	Female	TRUE
26	Sharvani	Female	i	0.180693069	0.81930693	Female	TRUE
27	Triouri	Female	i	0.180693069	0.81930693	Female	TRUE

Prediction Using F1 Score

- F1 Score – The harmonic mean between precision and recall which establishes the accuracy of the model in a much better manner.

$$2 * \text{Precision} * \text{Recall} / (\text{Precision} + \text{Recall})$$

- Precision** – The ratio of true values to our predicted true values ($\text{TP}/(\text{TP}+\text{FP})$)
- Recall** – The ratio of true values to the actual number of true values ($\text{TP}/(\text{TP}+\text{FN})$)
- To depict the values, we create a table called **Confusion matrix**



CountOfGender	Category			
Gender	Female	Male	(blank)	Grand Total
Female	367	50		417
Male	125	358		483
(blank)				
Grand Total	492	408		900
Precision		75%		
Recall		88%		
F1 SCORE=		81%		

Testing the Model On Random Names-:

After building the model, tested it on random names using:

- **VLOOKUP()** - Find the most probable gender based on the last letter.
- **IF()** - Compares probabilities and assigns gender.

	A	B	C	D	E
1	Name ▼	LastLetter ▼	Male Chances ▼	Female Chances ▼	Gender ▼
2	Zaraaya	a	26.96%	73.04%	Female
3	Nirvaan	n	92.89%	7.11%	Male
4	Vihaana	a	26.96%	73.04%	Female
5	Saanvi	i	18.07%	81.93%	Female

Conclusion :

- Gender can be predicted using names with **81% accuracy** using Excel functions.
- **Pivot Tables & VLOOKUP** help identify patterns in last-letter frequencies.