DATAPLAY

Predicting Gender from Names: A Practical Excel Solution

Presented by : Radhika Dwivedi

Introduction

- Objective : Leverage data-driven gender prediction to enhance business strategies.
- Business Value : Helps in targeted marketing, customer segmentation, and personalized outreach.
- Dataset Overview : 3002 records with Name and Gender columns.

Approach

1. Stratified Data Splitting : Ensures balanced 80-20 train-test ratio.
2. Feature Engineering : Extract last letter of each name as a predictor.
3. Pivot Table Analysis : Computes gender distribution by last letter.
4. Prediction Model : Uses VLOOKUP to assign gender probabilities.

Model Implementation

- **Pivot Table:**
  - **Rows: Last Letter**
  - **Values: Percentage of gender distribution.**
- **Prediction Logic:**
  - **Male Probability: =VLOOKUP(C2,training!$H$1:$J$26,3,0)**
  - **Female Probability: =VLOOKUP(C2,training!$H$1:$J$26,2,0)**
- **Final Gender Classification: =IF(D3>E3,"Male","Female")**

Raw Data

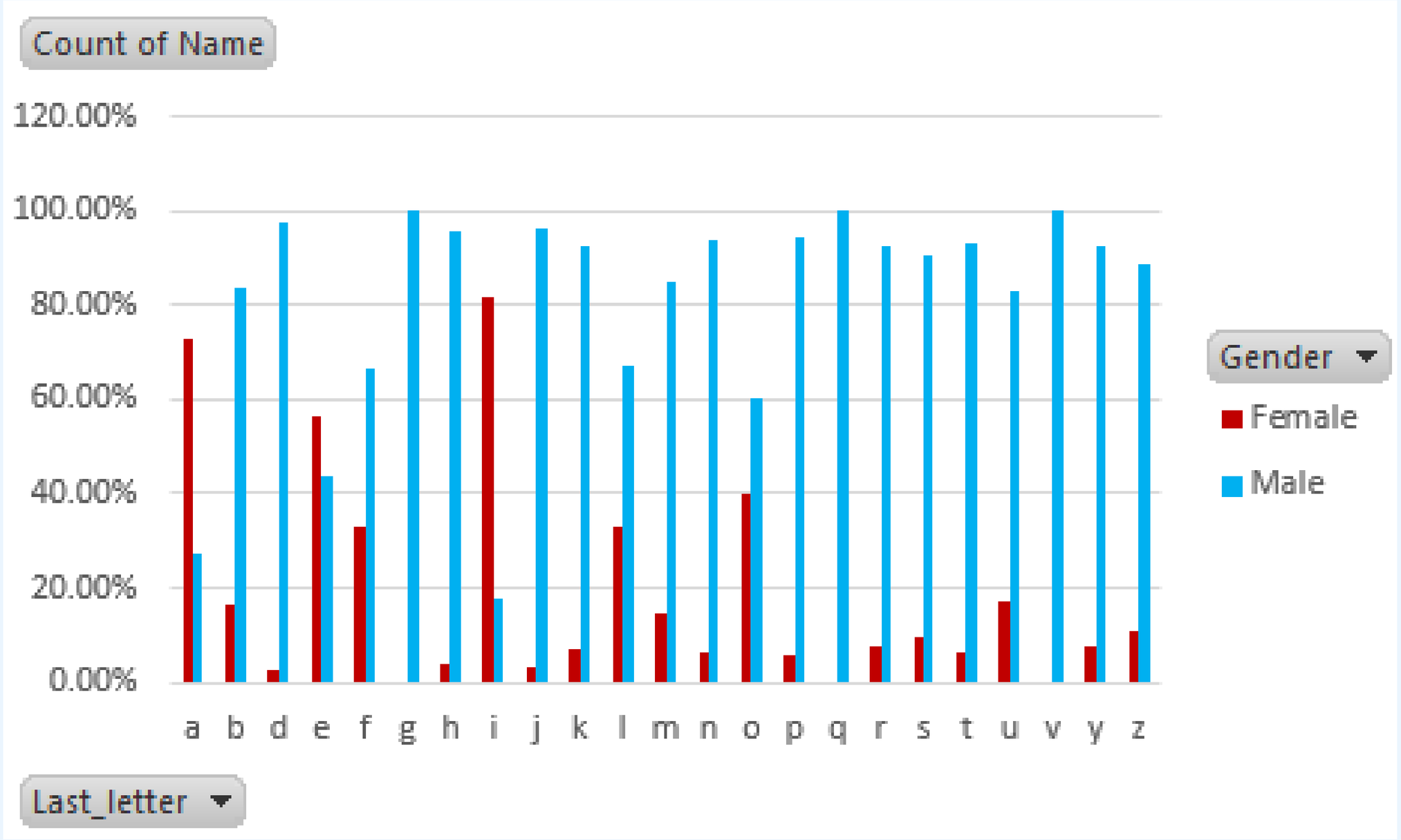
Name	Gender	LastLet
Ashutosh	Male	h
Meghama	Female	a
Sahib	Male	b
Pragya	Female	a
Kranti	Female	i
Tulika	Female	a
Aarushi	Female	i
Abhicand	Male	a
Pratigya	Female	a
Devak	Male	k
Kashipras	Male	d
Madhavi	Female	i
Charusila	Female	a
Chithayu	Male	u
Manmayi	Female	i
Mahajabe	Female	n
Krishnaku	Male	r
Kailas	Male	s
Nidhyath	Female	i

Model

Count of Name	Column Labels		
Row Labels	Female	Male	Grand Total
a	72.87%	27.13%	100.00%
b	16.67%	83.33%	100.00%
d	2.38%	97.62%	100.00%
e	56.25%	43.75%	100.00%
f	33.33%	66.67%	100.00%
g	0.00%	100.00%	100.00%
h	4.08%	95.92%	100.00%
i	81.92%	18.08%	100.00%
j	3.45%	96.55%	100.00%
k	7.27%	92.73%	100.00%
l	33.03%	66.97%	100.00%
m	14.81%	85.19%	100.00%
n	6.42%	93.58%	100.00%
o	40.00%	60.00%	100.00%
p	5.88%	94.12%	100.00%
q	0.00%	100.00%	100.00%
r	7.69%	92.31%	100.00%
s	9.68%	90.32%	100.00%
t	6.67%	93.33%	100.00%
u	17.24%	82.76%	100.00%
v	0.00%	100.00%	100.00%

Validation

Name	Gender	Last_letter	Gender_M	Gender_F	Gender_Pred	Correct/Incorrect
Garima	Female	a	0.27134503	0.728654971	Female	TRUE
Shambhav	Female	i	0.18082789	0.819172113	Female	TRUE
Sadiqua	Female	a	0.27134503	0.728654971	Female	TRUE
Devamati	Female	i	0.18082789	0.819172113	Female	TRUE
Inayat	Female	t	0.93333333	0.066666667	Male	FALSE
Namita	Female	a	0.27134503	0.728654971	Female	TRUE
Hariganga	Female	a	0.27134503	0.728654971	Female	TRUE
Manjira	Female	a	0.27134503	0.728654971	Female	TRUE
Mudrika	Female	a	0.27134503	0.728654971	Female	TRUE
Harshitha	Female	a	0.27134503	0.728654971	Female	TRUE
Nalini	Female	i	0.18082789	0.819172113	Female	TRUE
Odathi	Female	i	0.18082789	0.819172113	Female	TRUE
Soumya	Female	a	0.27134503	0.728654971	Female	TRUE
Sunayana	Female	a	0.27134503	0.728654971	Female	TRUE
Rishika	Female	a	0.27134503	0.728654971	Female	TRUE
Geena	Female	a	0.27134503	0.728654971	Female	TRUE
Esha	Female	a	0.27134503	0.728654971	Female	TRUE
Apsara	Female	a	0.27134503	0.728654971	Female	TRUE
Sarada	Female	a	0.27134503	0.728654971	Female	TRUE
Chaitali	Female	i	0.18082789	0.819172113	Female	TRUE
Sevati	Female	i	0.18082789	0.819172113	Female	TRUE
Jehannaz	Female	z	0.88888889	0.111111111	Male	FALSE

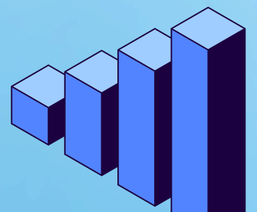

Count of Name
120.00%
100.00%
80.00%
60.00%
40.00%
20.00%
0.00%
a b d e f g h i j k l m n o p q r s t u v y z
Gender
Female
Male
Last_letter

Business Impact

- Accuracy Achieved: 83%
- Use Cases:
  - Personalized Marketing: Gender-specific campaigns.
  - Customer Insights: Better demographic understanding.
  - Lead Segmentation: Optimized outreach strategies.

Conclusion & Future Scope

- Simple yet effective gender prediction using Excel.
- Enables data-driven decision-making for business growth.



Next Steps:

- Enhance with machine learning for higher accuracy.
- Integrate with CRM systems for automated insights.
- Expand features beyond last-letter analysis for better precision.

Thank You