

# Medical Insurance Cost Prediction

presentation by:-Anjali Singh  
Shivankar  
Rohit Singh

age	sex	region	urban_rural	income	education	marital_status	
22	52	Female	North	Suburban	22700	Doctorate	Married
35	79	Female	North	Urban	12800	No HS	Married
55	68	Male	North	Rural	40700	HS	Married
00	15	Male	North	Suburban	15600	Some College	Married
92	53	Male	Central	Suburban	89600	Doctorate	Married
35	63	Female	North	Rural	305000	HS	Single
05	36	Male	West	Rural	38900	Masters	Single
18	21	Female	South	Suburban	83700	HS	Single
58	53	Male	North	Suburban	60700	No HS	Divorced
75	28	Male	South	Urban	23600	Bachelors	Married
02	34	Female	South	Urban	98900	Masters	Divorced
42	69	Female	Central	Suburban	19800	HS	Married
31	50	Male	East	Rural	43100	Masters	Married
39	33	Male	North	Suburban	65000	Bachelors	Married
25	63	Male	South	Urban	100700	Bachelors	Single
03	52	Male	South	Suburban	31900	Some College	Married
50	52	Male	West	Urban	17700	Bachelors	Single
52	59	Male	East	Urban	61900	Masters	Married
21	24	Male	East	Rural	158700	HS	Divorced
52	53	Male	Central	Urban	41300	Bachelors	Single
34	54	Female	South	Suburban	76400	Bachelors	Married
96	61	Female	South	Urban	211200	Masters	Single
97	54	Male	South	Rural	6600	Masters	Single
18	32	Female	North	Suburban	16000	Some College	Married
32	42	Female	North	Rural	26900	Bachelors	Married
39	68	Female	East	Urban	69500	Bachelors	Married
37	43	Female	East	Urban	15400	Bachelors	Single
93	43	Female	West	Urban	47600	Doctorate	Married
32	45	Female	West	Urban	201300	Doctorate	Single
31	68	Male	West	Suburban	36100	HS	Married
23	39	Other	South	Suburban	38800	Bachelors	Single
37	76	Male	East	Urban	4900	Masters	Single
22	32	Male	East	Suburban	41200	HS	Married
34	60	Male	North	Urban	23500	Some College	Married
03	82	Female	West	Suburban	26000	HS	Divorced
34	56	Female	Central	Urban	14200	Masters	Widowed
70	63	Female	West	Urban	49200	HS	Single
06	57	Female	East	Urban	38100	Masters	Married
3	40	Female	Central	Urban	16600	Bachelors	Married

# About Dataset

This dataset provides information about 100,000 individuals including their demographics, socioeconomic status, health conditions, lifestyle factors, insurance plans, and medical expenditures.

# Data Preprocessing



## Handle Outliers (Data Audit)

Cap extreme values (e.g., incomes, costs, or BMI) using winsorization or percentile capping.

## Encode Categorical Variables (Type)

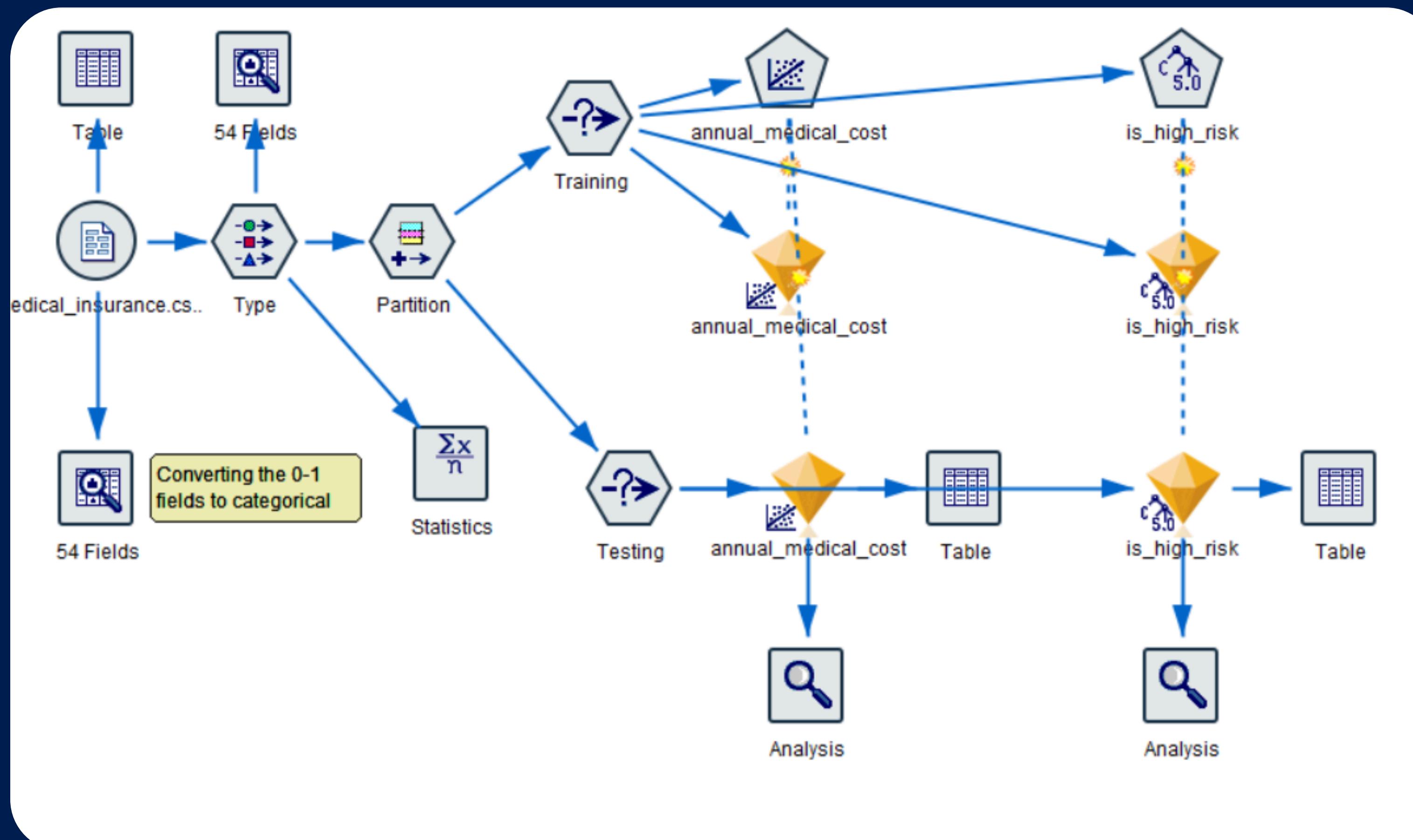
Convert sex, region, plan\_type, etc. to nominal.

## Split Data

Use Partition Node to split:

- 70% for training
- 30% for testing





# Annual Medical Cost

Metric	Meaning	Interpretation
Minimum Error	Smallest prediction error ( $\text{Predicted} - \text{Actual}$ )	-11,251 means one prediction underestimated cost by ₹11k (normal for large-cost data).
Maximum Error	Largest positive error	9,895 means one overestimation by ₹9.8k.
Mean Error	Average signed error	1.15 → almost unbiased (model not systematically under/over-predicting). <span style="color: green;">✓</span>
Mean Absolute Error (MAE)	Average absolute deviation between predicted and actual	455.78 → on average, predictions are off by about ₹456. Very good given large range of medical costs. <span style="color: green;">✓</span>
Standard Deviation (of errors)	Variation in prediction errors	769.53 → moderate variability, expected.
Linear Correlation ( $r$ )	Correlation between predicted and actual values	0.97 → excellent correlation → <b>very strong linear relationship</b> <span style="color: green;">✓</span>

# High Risk Prediction

## Overview Table

Metric	Meaning	Result
Correct	Total correct predictions	$29,927 / 29,928 = 99.996\%$ accuracy 
Wrong	Total misclassified cases	Only 1 wrong → almost perfect prediction.
Total	Total test records	29,928

## Coincidence Matrix (Confusion Matrix)

Actual (Row) \ Predicted (Column)	Predicted 0	Predicted 1
Actual 0	18,908	0
Actual 1	1	11,019

- 18,908 **true negatives** (correctly predicted low-risk)
- 11,019 **true positives** (correctly predicted high-risk)
- Only 1 **false negative** (a high-risk case predicted as low-risk)
- **No false positives**

A photograph showing two people in business attire shaking hands over a desk. On the desk, there is a laptop, a smartphone, and some papers. The background is slightly blurred.

# The End