



**INDIANA UNIVERSITY**  
BLOOMINGTON

INDIANA UNIVERSITY BICENTENNIAL

# Exploratory Data Analysis of Mobile Price Classification Data

## Team Members

Ashish Patidar  
Rajesh Sharma



200 YEARS

INDIANA UNIVERSITY BICENTENNIAL

# Research Goals

- To find the component with the most impact on the price
- Build a multinomial logistic model for the price range prediction with the most impactful component as predictor
- Combine the predictor with other components for more accurate model



# Data Description

- 2000 rows
- 14 columns
  - 10 numerical
  - 4 categorical

## COLUMN NAMES

battery_power	px_width
clock_speed	px_height
pc (primary camera)	sc_h (screen height)
ram	sc_w (screen width)
int_memory	n_cores
four_g	touch_screen
dual_sim	price

Source: <https://www.kaggle.com/datasets/iabhishekofficial/mobile-price-classification>

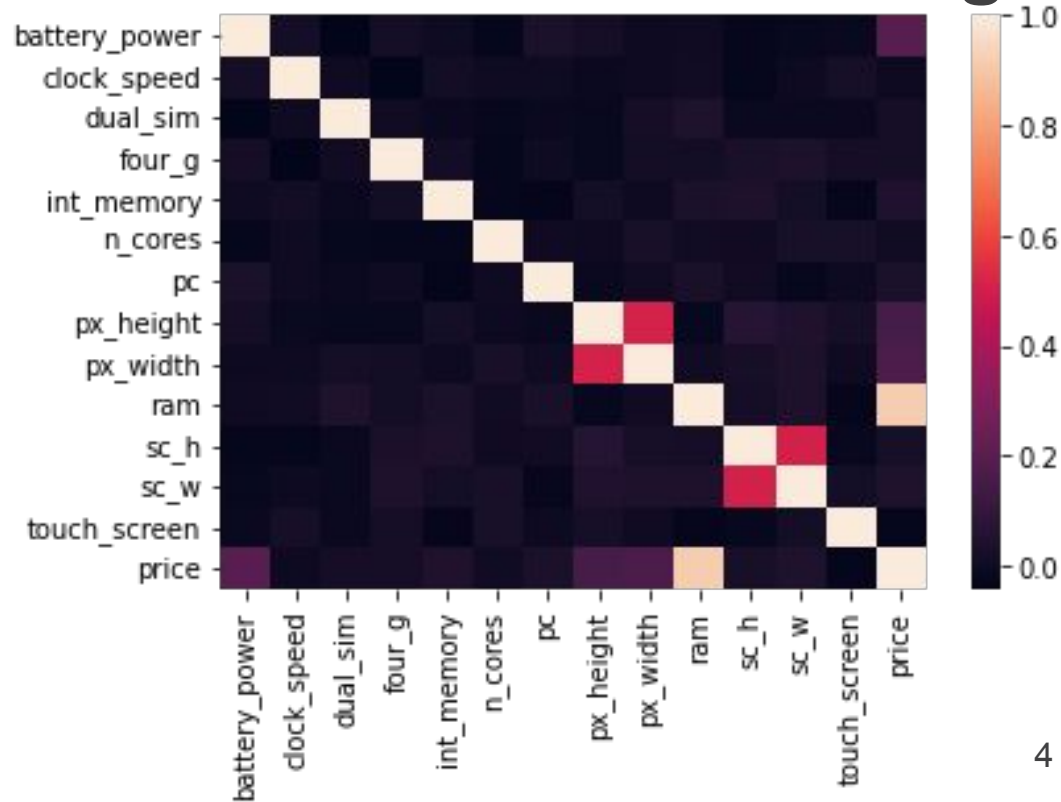


200 YEARS

INDIANA UNIVERSITY BICENTENNIAL

# Preliminary Analysis and Data Pre-Processing

- $\text{performance} = \text{clock\_speed} * \text{n\_cores}$
- $\text{px\_res} = \text{px\_height} * \text{px\_width}$
- $\text{sc\_length} = \sqrt{(\text{sc\_h})^2 + (\text{sc\_w})^2}$





200 YEARS

INDIANA UNIVERSITY BICENTENNIAL

# Data Snapshot

	battery_power	dual_sim	four_g	int_memory	pc	ram	touch_screen	price	performance	px_res	sc_length
0	842	0	0	7	2	2549	0	Medium	4.4	15120	11.401754
1	1021	1	1	53	6	2631	1	High	1.5	1799140	17.262677
2	563	1	1	41	6	2603	1	High	2.5	2167308	11.180340
3	615	0	0	10	9	2769	0	High	15.0	2171776	17.888544
4	1821	0	1	44	14	1411	1	Medium	2.4	1464096	8.246211

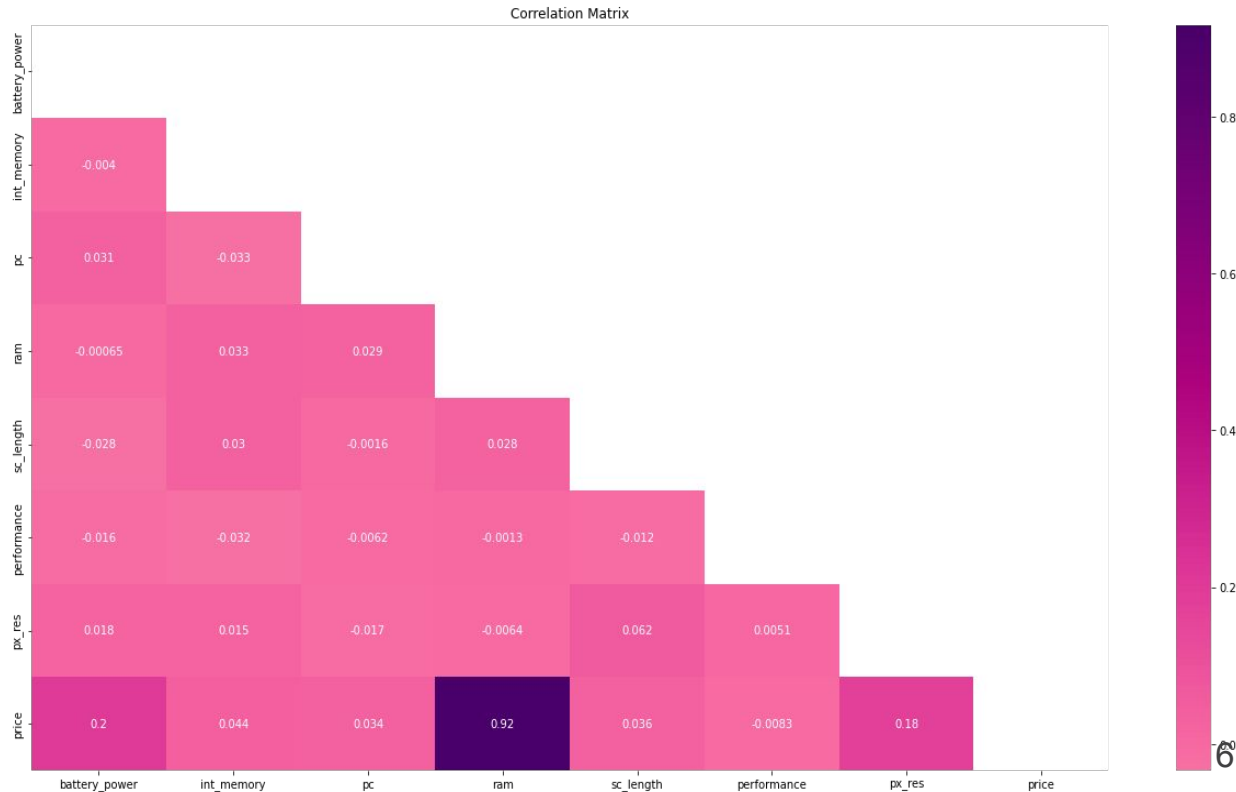


200 YEARS

INDIANA UNIVERSITY BICENTENNIAL

# CORRELATION PLOT

- Ram is only predictor variable which is highly correlated with response variable price
- Very little or almost no correlation between other predictor variable
- Battery\_power and px\_res (pixel resolution) can be interesting variable to combine with ram





200 YEARS

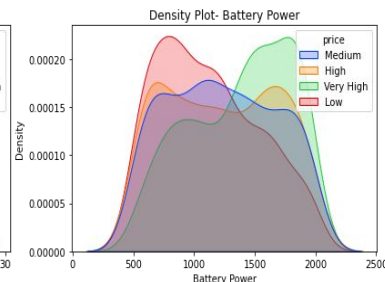
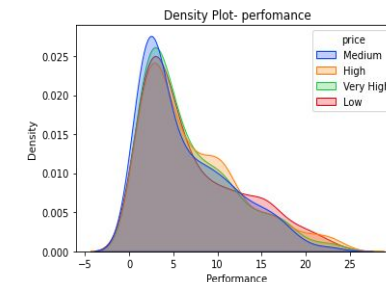
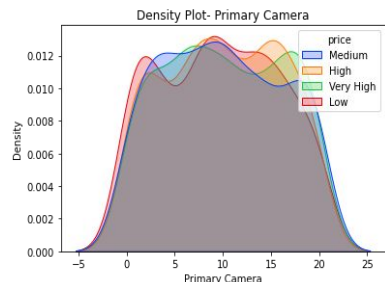
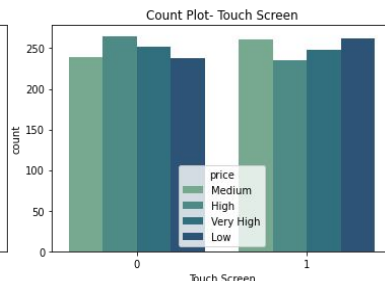
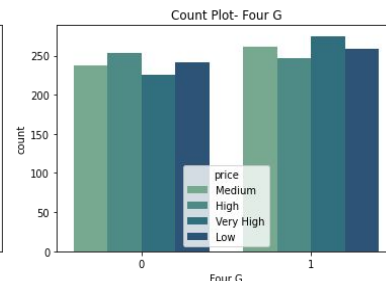
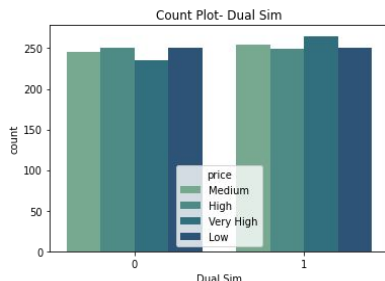
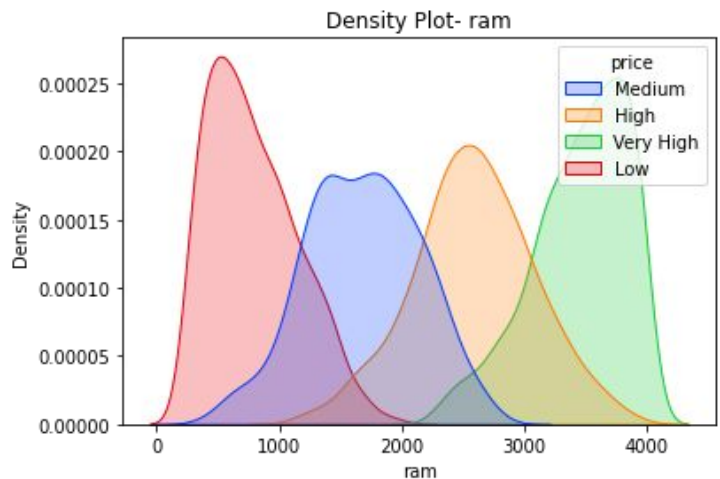
INDIANA UNIVERSITY BICENTENNIAL

# UNIVARIATE ANALYSIS

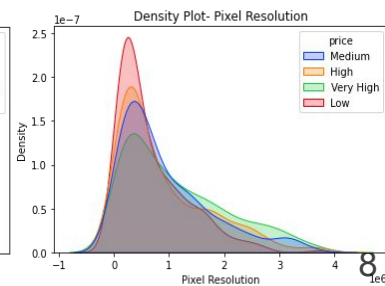
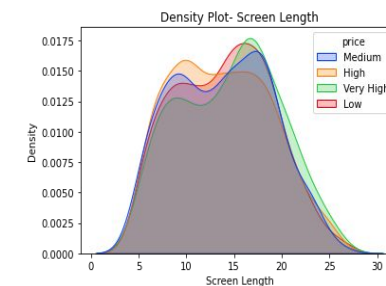
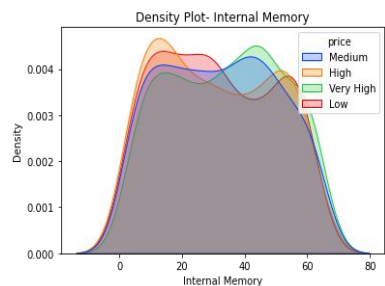


200 YEARS

INDIANA UNIVERSITY BICENTENNIAL



- Visualization supports our intuition of RAM being the primary variable
- Distribution of other variables is same for all the price ranges







200 YEARS

INDIANA UNIVERSITY BICENTENNIAL

# BIVARIATE ANALYSIS

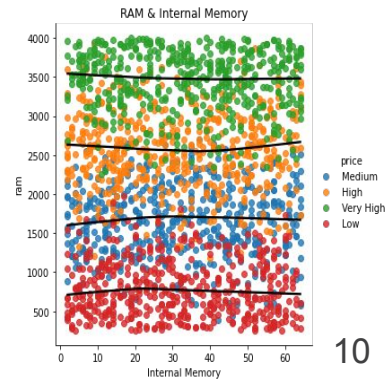
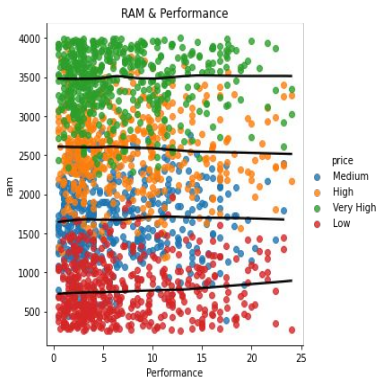
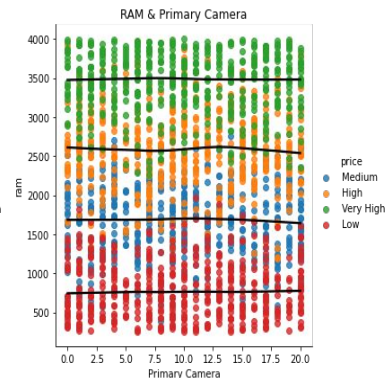
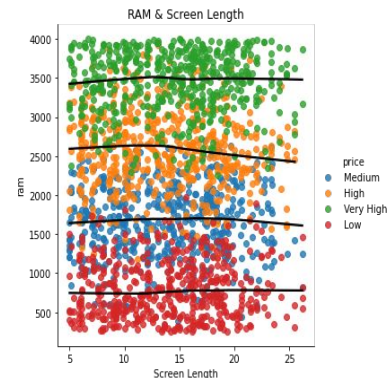
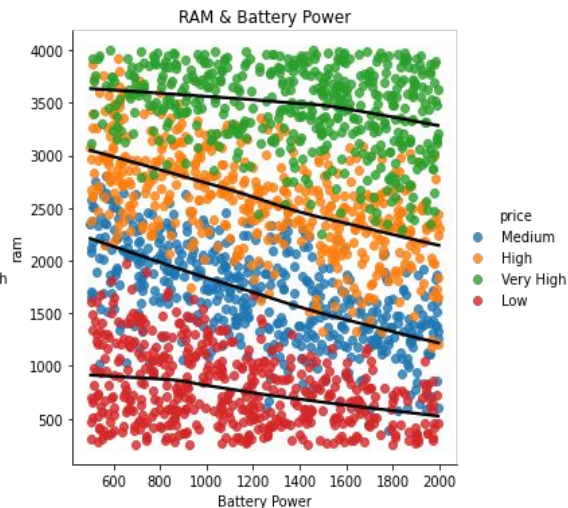
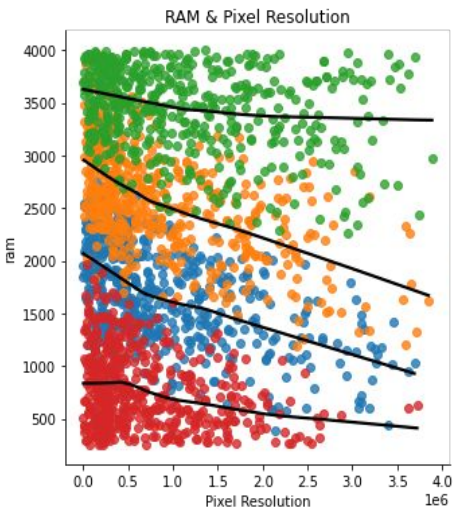


200 YEARS

INDIANA UNIVERSITY BICENTENNIAL

# RAM is the only important feature?

- Combining RAM with battery power or pixel resolution giving some interesting insights
- Money can get you anything





# Multinomial logistic models

- Started with proportional odds logistic model but since we have same number of phones in each category, it cannot be implemented
- Used nnet package of R to implement multinomial logistic regression

```
ram.mlog = nnet :: multinom(price_range ~ ram + px_res + battery_power,  
                             data = mobile_data)
```



# Model Accuracy

PREDICTOR(S)	RESPONSE	ACCURACY
Ram	Price Range	72.5%
Ram + Battery Power	Price Range	82.5%
Ram + Pixel Resolution	Price Range	80.2%
Ram + Battery Power + Pixel Resolution	Price Range	93.2%



# Interpretation of Analysis

- Ram seems to be the most important predictor
- Multinomial logistic regression is best model
- Individually Battery Power and Pixel Resolution cannot be used as predictors for classification but combining them with RAM increases the model accuracy by a good margin
- Can use ML techniques for better prediction, increasing complexity of model



# Limitations

- Chances of overfitting (as testing on same dataset)
- Cannot fit proportional odds logistic model, which reduces the interpretability of model
- Artificial dataset, 20 very high priced phones does not have wifi (Strange), so we cannot relate it with real world mobile phones



200 YEARS

INDIANA UNIVERSITY BICENTENNIAL

# Thank You