DISCRIMINATORY PRICING

FEATURES (410 Rows X 24 Columns)

psoda: price of medium soda

pfries: price of small fries

pentree: price entree

pmeal: price meal

~~wagest: starting wage~~

~~nmgrs: number of managers~~

~~nregs: number of registers~~

~~hrsopen: hours open~~

~~emp: number of employees~~

~~compown: =1 if company owned~~

~~chain: BK = 1, KFC = 2, Roy Rogers = 3,~~
~~Wendy's = 4~~

density: population density, town

crmrte: crime rate, town

state: NJ = 1, PA = 2

prpblck: proportion black, zipcode

prppov: proportion in poverty, zipcode

prpncar: proportion no car, zipcode

~~hseval: median housing value, zipcode~~

~~nstores: number of stores, zipcode~~

income: median family income, zipcode

county: county label

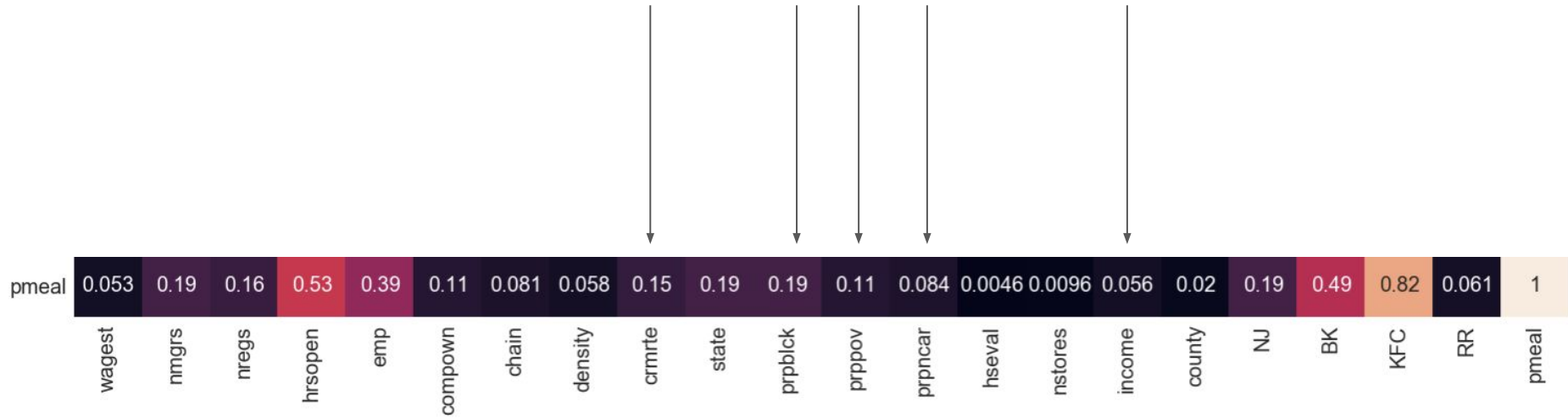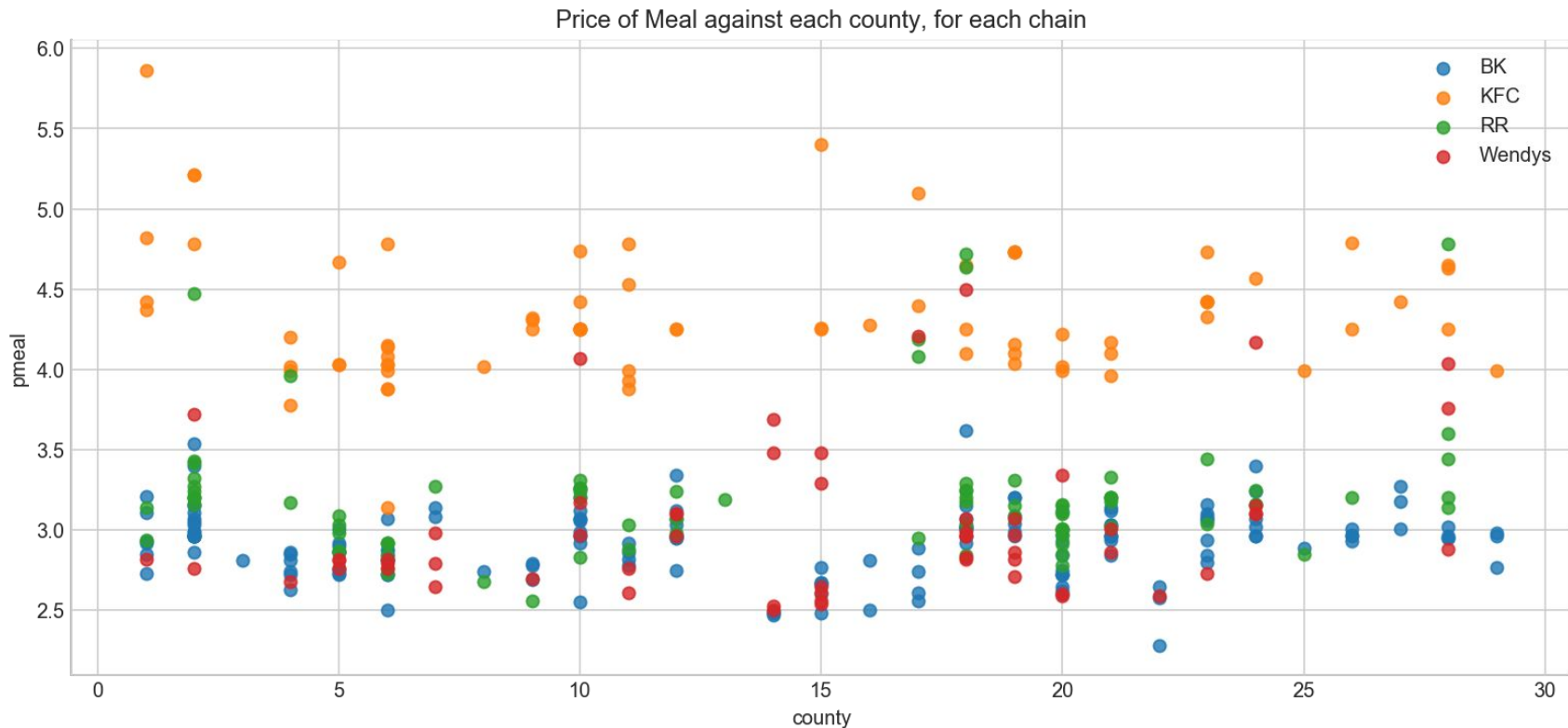NJ=1 for New Jersey

BK=1 if Burger King
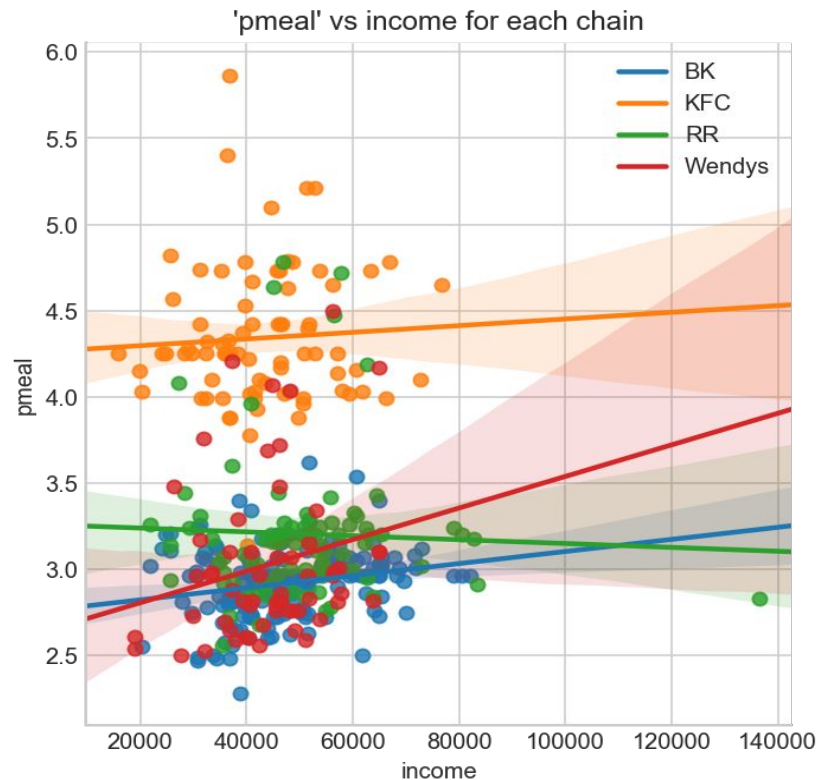
KFC=1 if Kentucky Fried Chicken

RR=1 if Roy Rogers

EDA

	wagest	nmgrs	nregs	hrsopen	emp	compown	chain	density	crmrte	state	prpblck	prppov	prpncar	hseval	nstores	income	county	NJ	BK	KFC	RR	pmeal
pmeal	0.053	0.19	0.16	0.53	0.39	0.11	0.081	0.058	0.15	0.19	0.19	0.11	0.084	0.0046	0.0096	0.056	0.02	0.19	0.49	0.82	0.061	1

EDA – Diff 'brands' have diff pricings across counties

Price of Meal against each county, for each chain

EDA
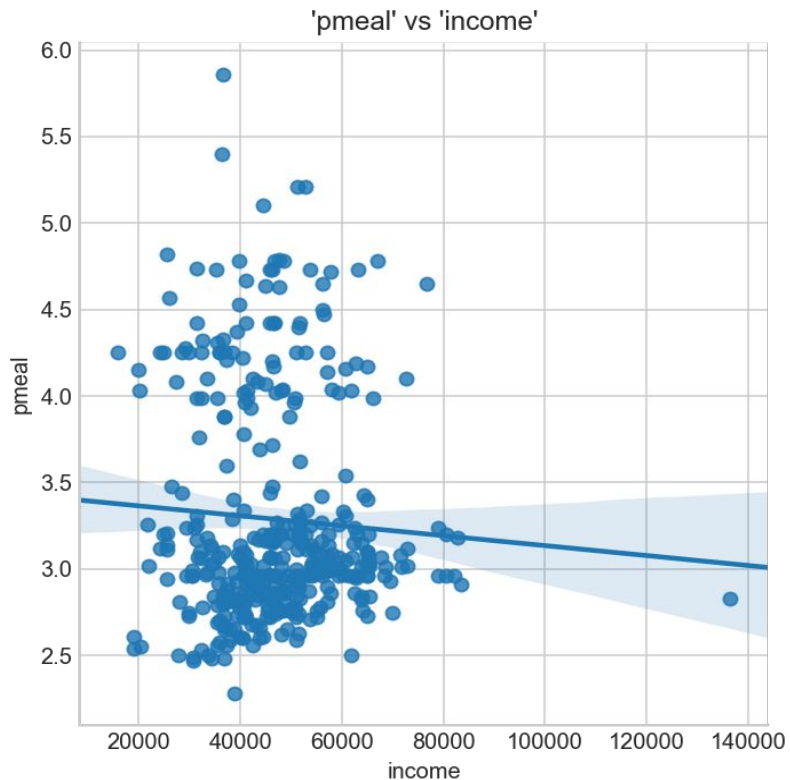
T-test

- H-null: The mean difference of the 'pmeal' of each group is 0.
- H-alt: "   "   is not 0.
  → The chain is practicing price discrimination for that feature.

T-test

- Split data by chain
- For each feature: split into 2 groups by the median

BK

Income <= 46362 (Median of Income)	**Income >** 46362 (Median of Income)
'pmeal' DATA	*'pmeal' DATA*

T-test Result

BK:	Income <= 46362 (Median of Income)	Income > 46362 (Median of Income)

PVALUE = **0.00278**

Reject null hypothesis, BK practice price discrimination due to income

Wendy's:	Crime Rate <= 0.0482 (Median of Crime Rate)	Crime Rate => 0.0482 (Median of Crime Rate)

PVALUE = **0.0235**

Reject null hypothesis, Wendy's practice price discrimination due to crime rate

Regression – ScikitLearn – LinearRegression()

predictors= ['KFC' ,'RR' ,'BK' ,'income' ,'crmrte']

Running single test - R^2 score: 0.7148012472090188

Cross-Validation (cv = 10) - Mean of R^2 Score: 0.6731695730264797

- Std deviation of R^2 Score: 0.08275667233214133

Regression – Statsmodel

- R^2 score: 0.961

	coef	std err	t	P>|t|	[0.025	0.975]
KFC	1.9718	0.108	18.262	0.000	1.760	2.184
BK	0.4664	0.095	4.900	0.000	0.279	0.654
RR	0.5831	0.111	5.243	0.000	0.364	0.802
income	4.324e-05	1.57e-06	27.459	0.000	4.01e-05	4.63e-05
crmrte	7.9129	0.682	11.600	0.000	6.572	9.254

Conclusion

Income and crime rate affects the price but only for Burger King and Wendy's respectively.

What else:

- Standardize our variables