



Predicting Income Levels

Francesca Schott

General Assembly
Part Time Data Science
Fall 2016

–
**How much money does
someone make** based on
if they **exercised** today,
have a certain **job**, or where
they buy **groceries**?



Bureau of Labor Statistics American Time Use Survey (ATUS)



ATUS: Eating & Health Module



Intro to Data

→ **Kaggle**

Data set originally found on Kaggle.

→ **BLS - ATUS**

Yearly survey of how Americans spend their time during daily activities.

→ **ATUS - Eating & Health Module**

Additional questions about eating, meal preparation and health.



Shape
(11212, 43)

Data Subset

Reduced the rows to focus on the following subset of respondents:

- Single-Person Household
- Employed @ Work
- Eating & Health Respondent
- 2014 Poverty Threshold



Shape
(1158, 43)

Response Vars.

→ **Monthly Income Level (Categorical)**

Income levels based on poverty threshold set by US Census Bureau.

→ **Weekly Income (Continuous)**

Calculated based on respondent answers.



Shape

(1150, 43)

Monthly Income Level

Above 185% of
Poverty Threshold

Income > \$1,900

(869/1150)

Between 130% and
185% of Poverty
Threshold

\$1,300
< Income <
\$1,900

(145/1150)

Below 130% of
Poverty Threshold

Income < \$1,300

(136/1150)

Features

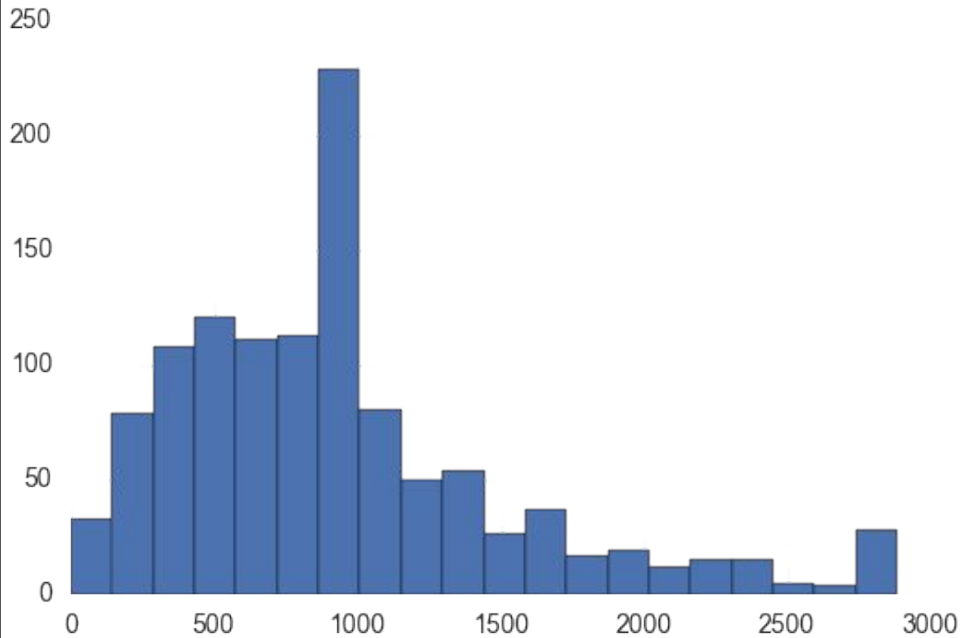
- Body Mass Index
- Primary Eating Time
- Secondary Eating Time
- Exercise
- Fast Food Purchases
- Food Amount
- Stores
- Occupation
- Occupation Industry



Shape
(1150, 11)

Histogram

Weekly Income

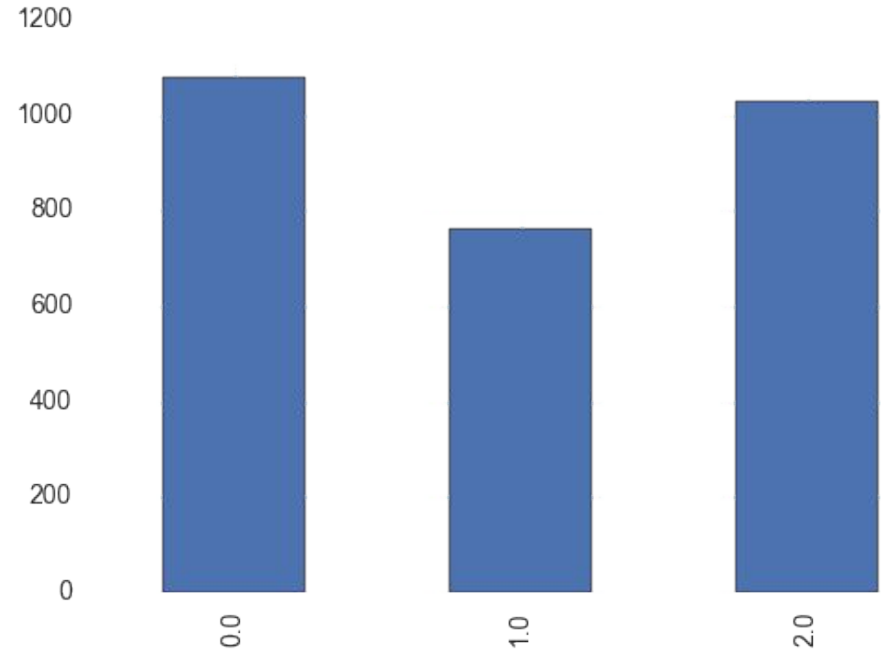


Bar Chart

—

X = Exercise

Y = Weekly Income



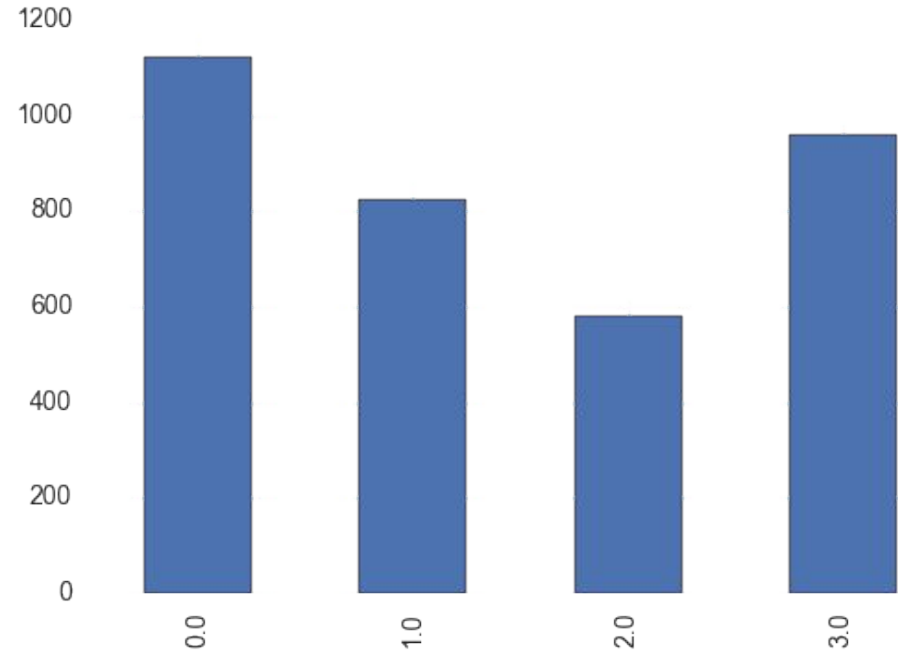
0 = Did not answer (5/1150)
1 = Did not exercise (378/1150)
2 = Did exercise (775/1150)

Bar Chart

—

X = Food Amount

Y = Weekly Income



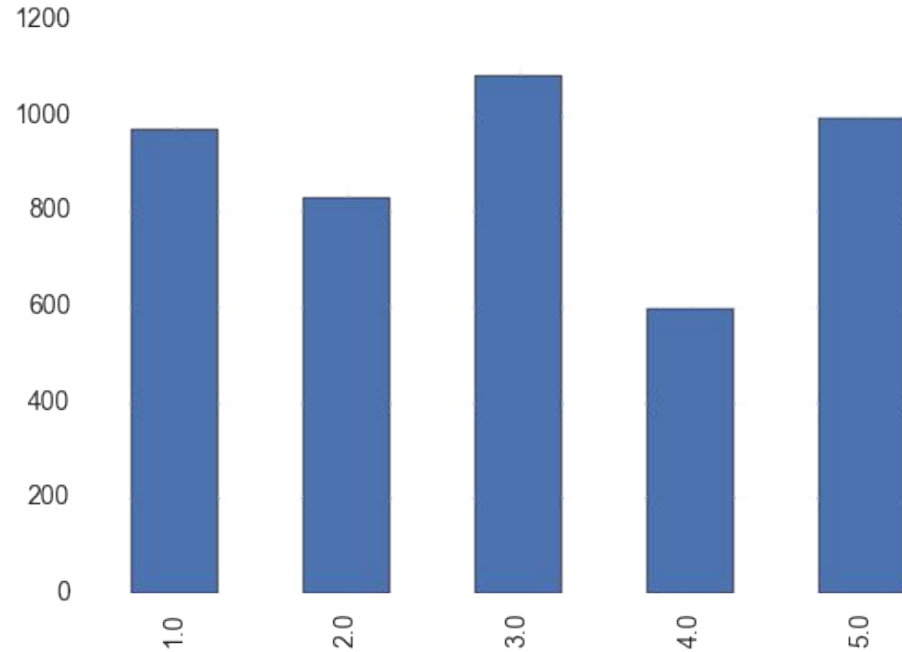
0 = Did not answer (6/1150)
1 = Often not enough to eat (17/1150)
2 = Sometimes not enough to eat (46/1150)
3 = Enough to eat (1089/1150)

Bar Chart

—

X = Stores

Y = Weekly Income



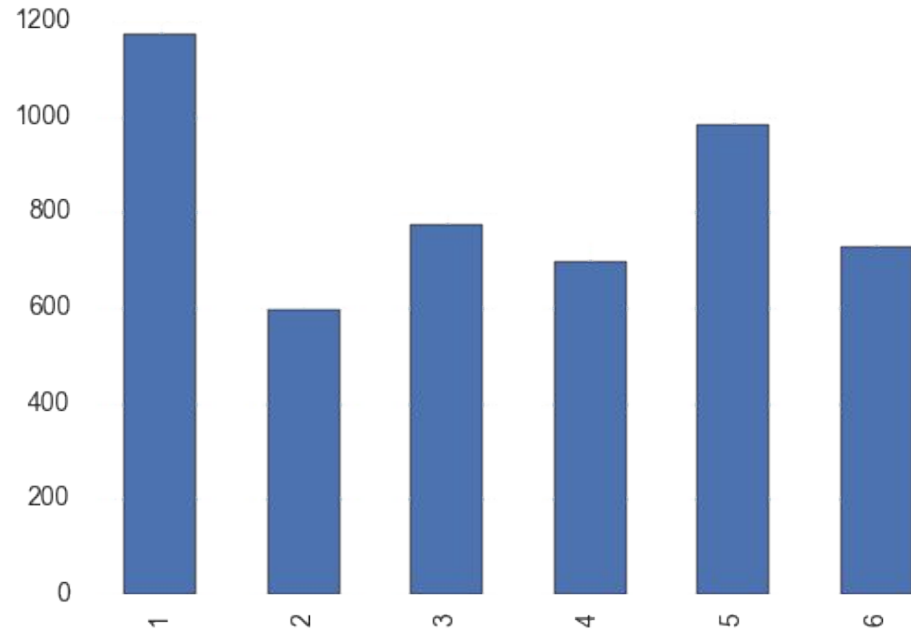
1 = Grocery store (832/1150)
2 = Supercenter (251/1150)
3 = Warehouse club (51/1150)
4 = Drugstore or convenience store (19/1150)
5 = Some other place (5/1150)

Bar Chart

—

X = Occupation
Category

Y = Weekly Income



1 = Management and professional	(543/1150)
2 = Service	(238/1150)
3 = Sales and office	(172/1150)
4 = Farming, fishing, and forestry	(126/1150)
5 = Construction and maintenance	(72/1150)
6 = Production, transportation, and material moving	(7/1150)

Categorical Modeling

Select Features

Exercise
Food Amount
Stores
Occupation Cat.

Dummies

OneHotEncoder

Testing

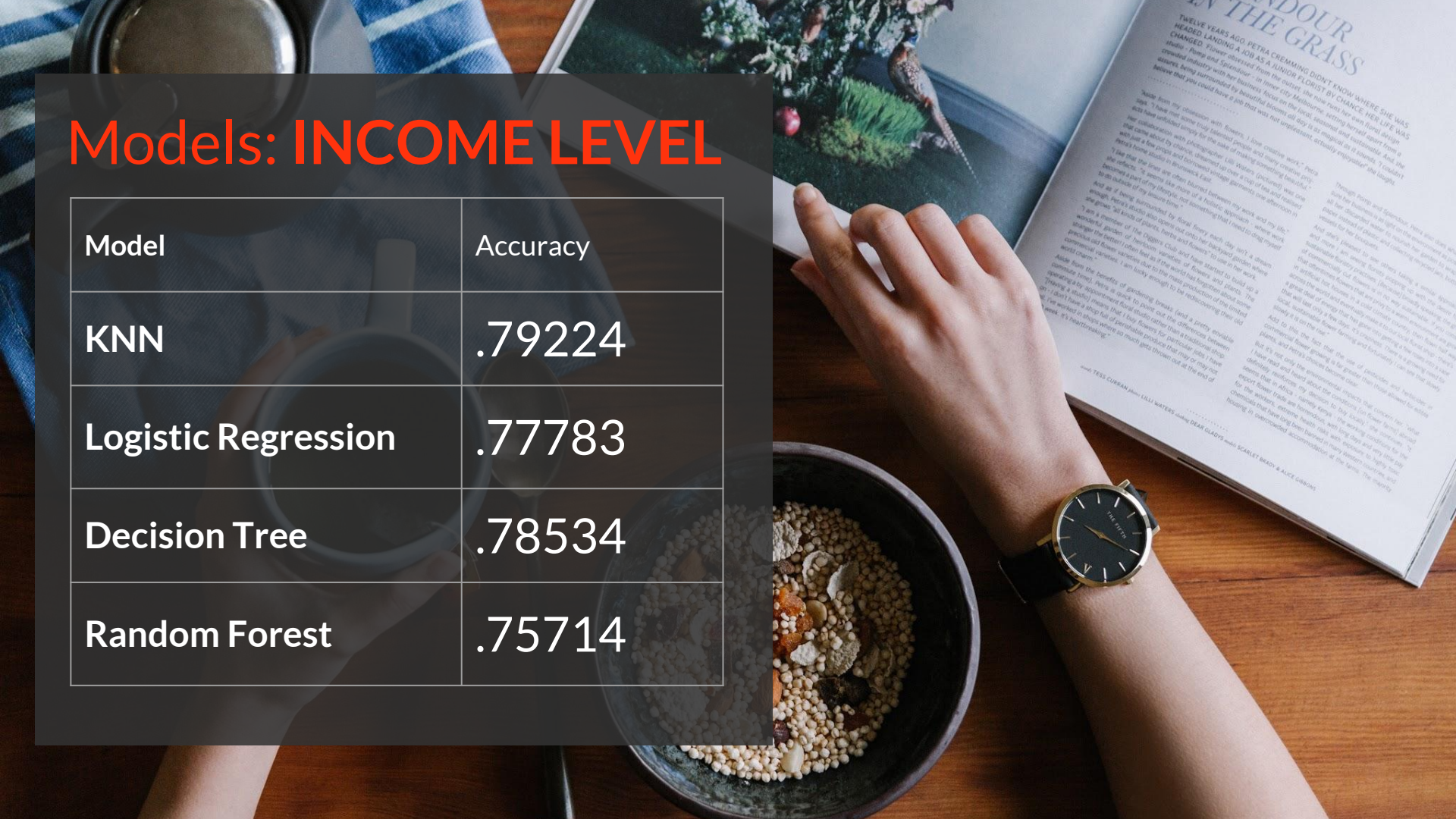
5-Fold
Cross-Validation

Null Accuracy

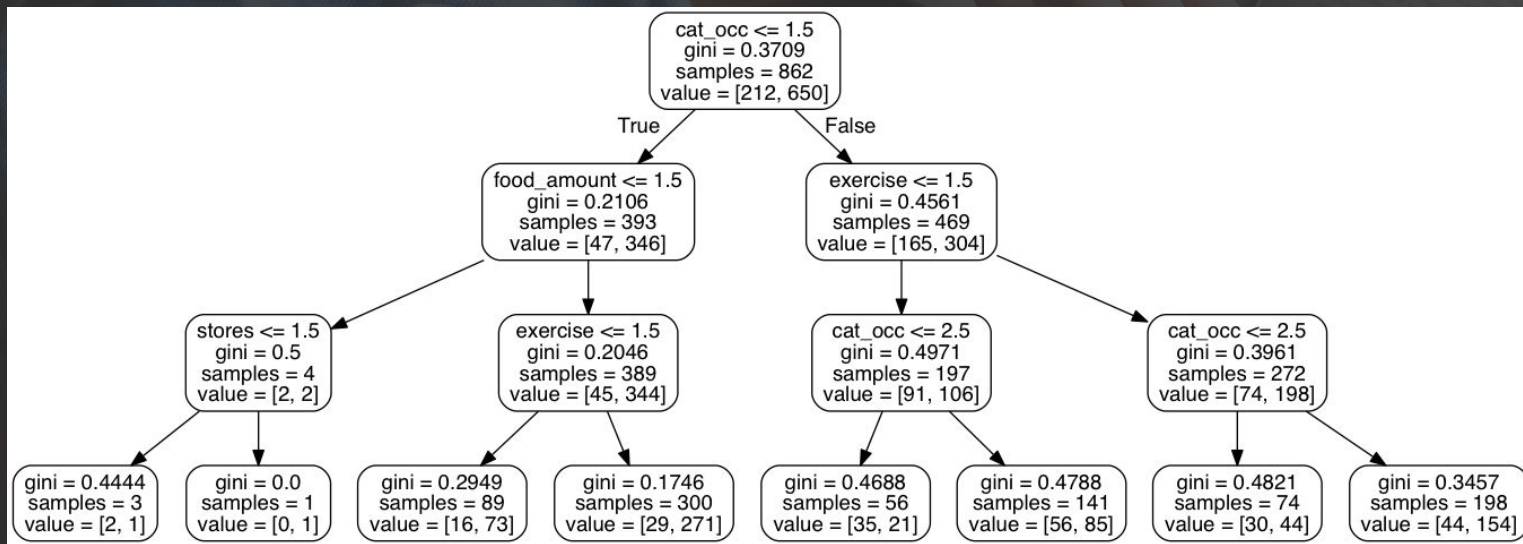
.75565

Models: INCOME LEVEL

Model	Accuracy
KNN	.79224
Logistic Regression	.77783
Decision Tree	.78534
Random Forest	.75714



Model: DECISION TREE



Model: DECISION TREE

Feature	Importance
Occupation Category	.73669
Exercise	.21944
Food Amount	.02794
Stores	.01593



Clustering WEEKLY INCOME

K-Means Clustering
Silhouette Coeff. = .57529



Clustering

WEEKLY INCOME

	Exercise	Food Amount	Store	Occupation Category	Weekly Income
0	2.000000	3.000000	1.228477	1.587748	1067.92363
1	1.572254	3.000000	1.358382	5.595376	826.925137
2	1.641509	2.981132	4.905660	2.566038	971.883505
3	0.992157	3.000000	1.258824	1.890196	782.221912
4	1.523077	1.646154	1.461538	2.707692	654.718915

Next Steps

→ **Improved Feature Selection**

→ **Parameter Tuning**

→ **Data Subset**

Multi-Person Household

2015 Poverty Threshold

→ **Other Features Beyond Income**

Many unused categories:

Health predictions

Welfare insights

Year-over-year trends

→ **Explore Weekly Income**

So...

To make more money...

- Exercise on the week the BLS calls for the ATUS
- Consider not shopping at convenience stores
- Work in management

Thank you

