$ Predicting Income Levels

Francesca Schott

General Assembly
Part Time Data Science
Fall 2016

How much money does
someone make

if they **exercised** today,
have a certain **occupation**,
or had **enough to eat**?

**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 levels set by US Census Bureau.

➔ **Weekly Income** (Continuous)
Calculated based on respondent answers.

Monthly Income Level

Above 185% of Poverty Threshold	Between 130% and 185% of Poverty Threshold	Below 130% of Poverty Threshold
Income > $1,900	**$1,300 < Income < $1,900**	**Income < $1,300**

Features

➔ **Body Mass Index**

➔ **Primary/Secondary Eating Time**

➔ **Exercise**

➔ **Fast Food Purchases**

➔ **Food Amount**

➔ **Stores**

➔ **Occupation**

➔ **Occupation Industry**

Shape
(1158, 11)

Histogram

—

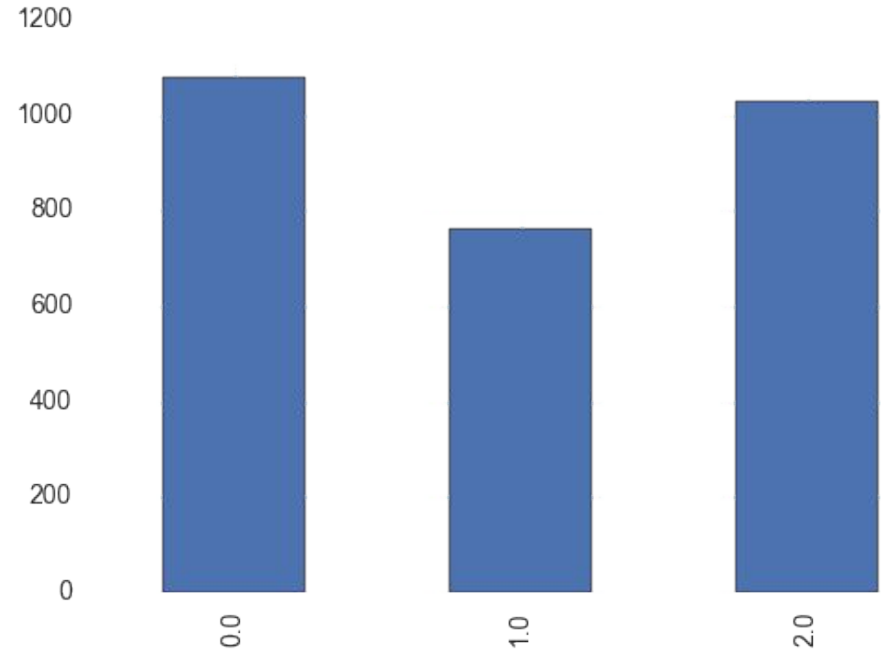# Weekly Income
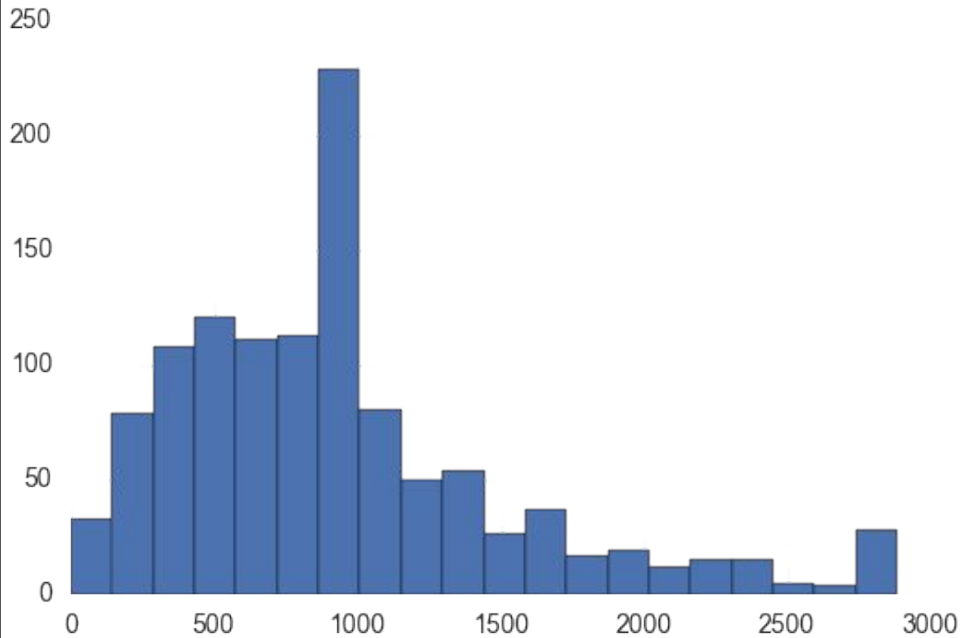
Bar Chart
—
X =   Exercise

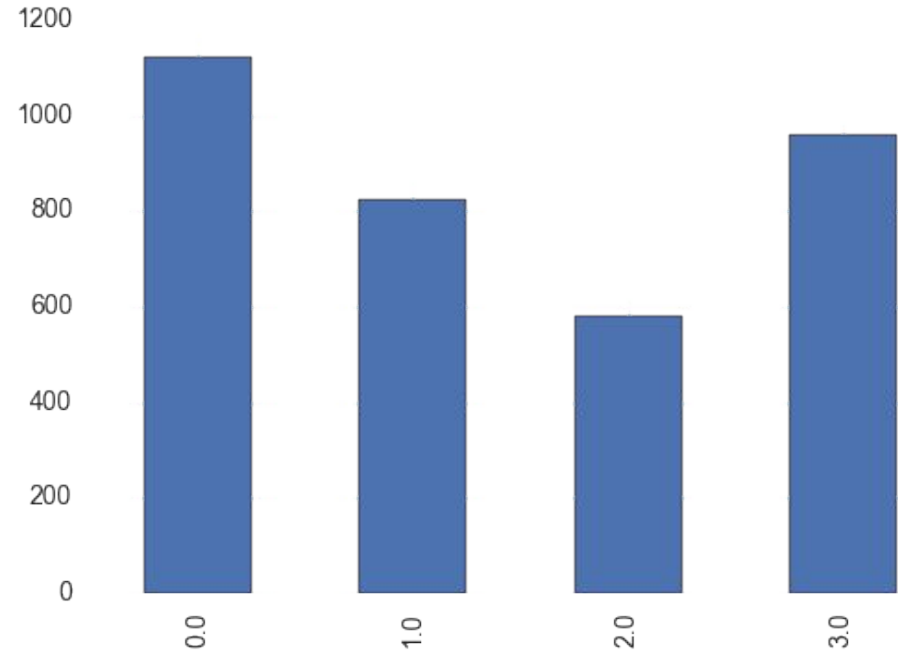Y =   Weekly Income

0 = Did not answer   (5/1158)
1 = Did not exercise   (378/1158)
2 = Did exercise   (775/1158)

Bar Chart

—

X =   Food Amount

Y =   Weekly Income

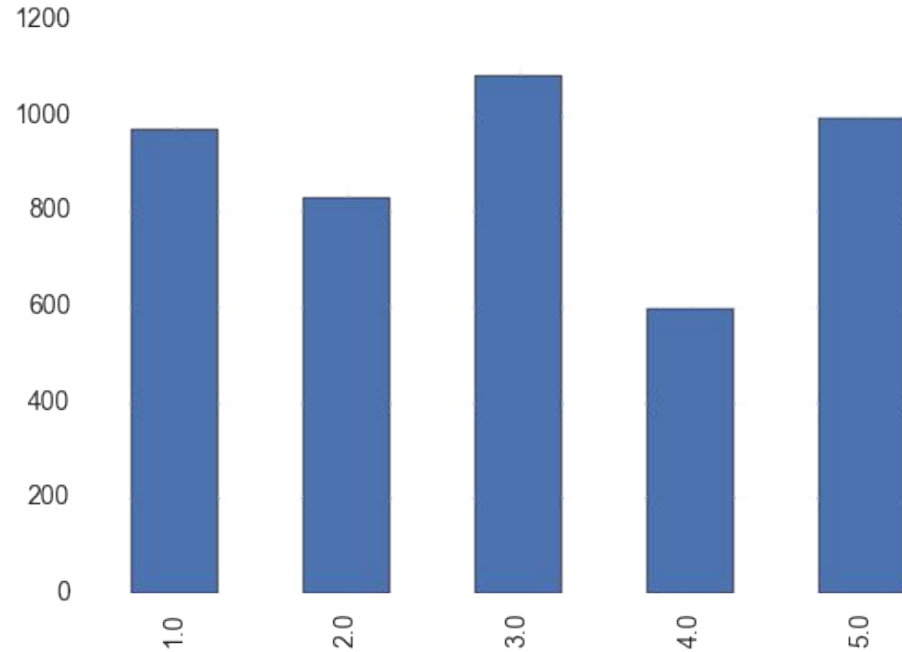0 = Did not answer   (6/1158)
1 = Often not enough to eat   (17/1158)
2 = Sometimes not enough to eat   (46/1158)
3 = Enough to eat   (1089/1158)

Bar Chart

—

X = Stores

Y = Weekly Income

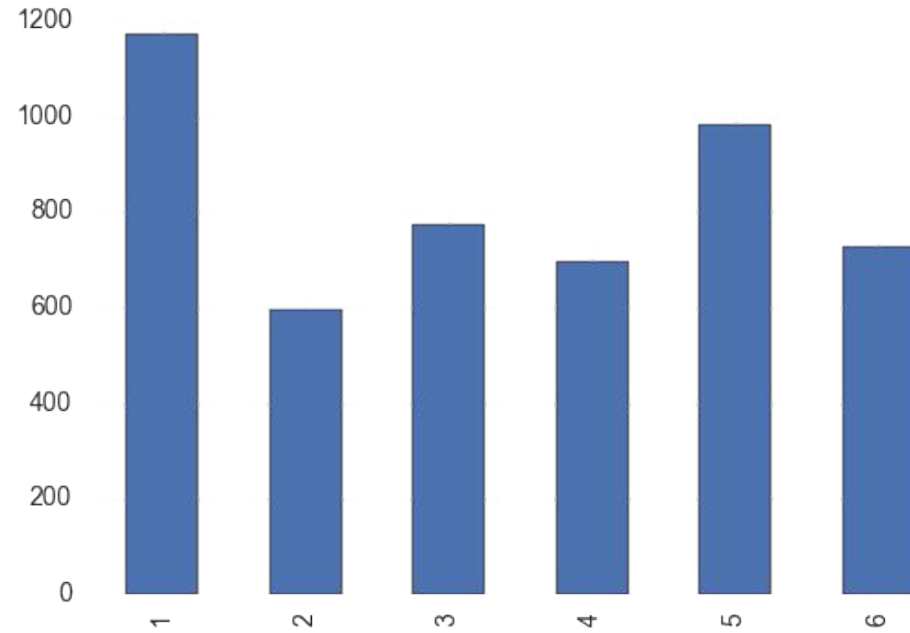1 = Grocery store   (832/1158)
2 = Supercenter   (251/1158)
3 = Warehouse club   (51/1158)
4 = Drugstore or convenience store(19/1158)
5 = Some other place   (5/1158)

Bar Chart

—

X =   Occupation Category

Y =   Weekly Income

1 = Management and professional   (543/1158)
2 = Service   (238/1158)
3 = Sales and office   (172/1158)
4 = Farming, fishing, and forestry   (126/1158)
5 = Construction and maintenance   (72/1158)
6 = Production, transportation,   (7/1158)
    and material moving

Categorical Modelling

Select Features

Exercise
Food Amount
Stores
Occupation Cat.

Models

KNN
Logistic Reg.
Decision Tree
Bagged Decision Tree
Random Forest
Multinomial NB

Testing

5-Fold
Cross-Validation

Models: INCOME LEVEL

KNN	.77616
Logistic Regression	.82114
Decision Tree	.79686
Random Forest	.79014

Models: INCOME LEVEL

Voting Classifier

Logistic Regression
Decision Tree
Random Forest

.80345

Clustering
WEEKLY INCOME

K-Means Clustering
Silhouette Coeff. = .56934

Clustering
WEEKLY INCOME

	Exercise	Food Amount	Store	Occupation Category	Weekly Income
0	1.703125	2.984375	4.578125	2.328125	1026.267710
1	1.577143	3.000000	1.360000	5.594286	828.378144
2	1.455882	1.573529	1.500000	2.735294	691.386437
3	2.000000	3.000000	1.194958	1.593277	1063.648576
4	0.992188	3.000000	1.257812	1.886719	783.014386

Next Steps

➔ **Data Subset**
Multi-Person Household
2014 Poverty Threshold

➔ **Other features?**
Many unused categories from ATUS and
EH

➔ **Explore Weekly Income**

➔ **Beyond Income**
Health predictions
Welfare insights
Year-over-year trends

So…

To make more money…

- Exercise on the day the BLS calls for the ATUS
- Don't shop at the convenience store on the week the BLS calls for the ATUS
- Work in management

Thank you 🙂