Customer Brand Preferences Report
Tasneem Dawoodjee

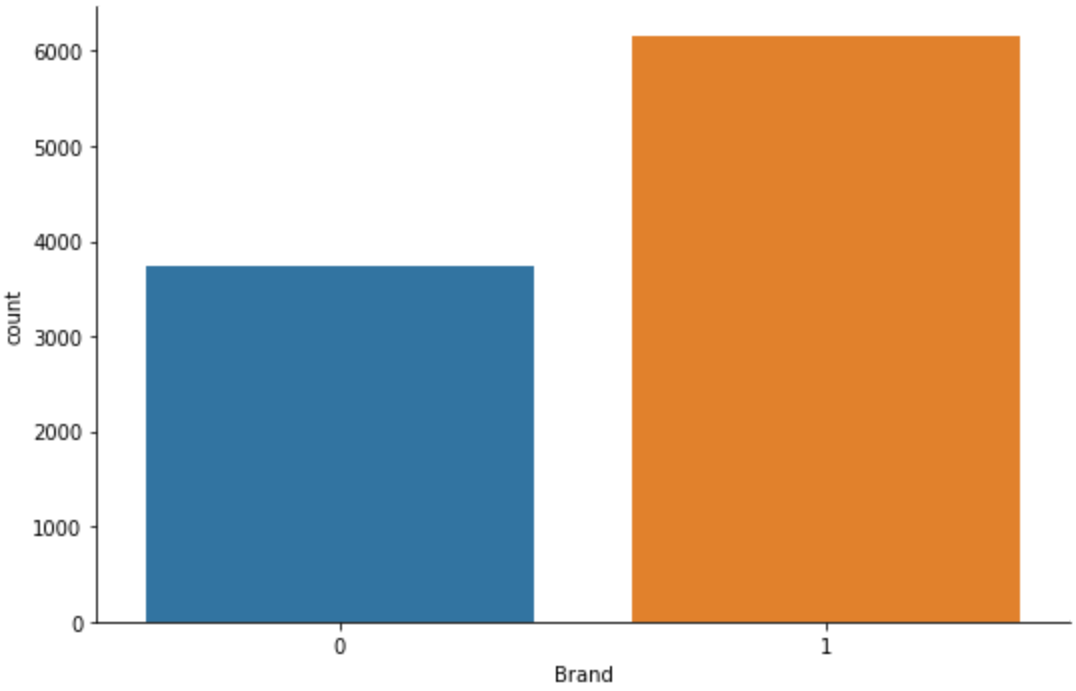
1. **Purpose**

The purpose of this project is to predict the customers' brand preferences that are missing from the incomplete surveys.

2. Variables

Brand

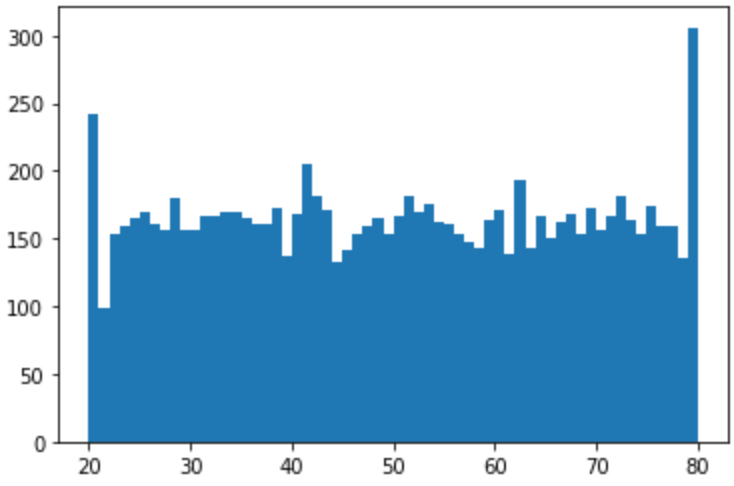
Figure 1: Barchart of Age



- Target variable
- 0 = Acer brand
- 1 = Sony brand
- Most customers prefer Sony
- Applied class imbalance techniques to address less preference for Acer brand

Age

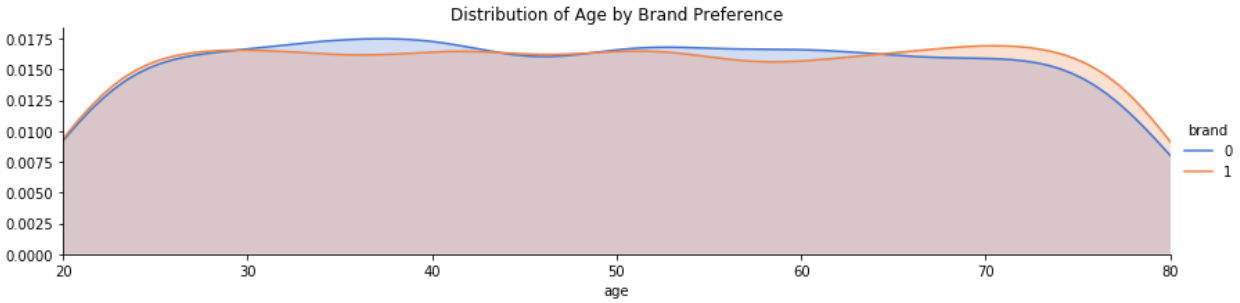
Figure 2: Histogram of Age



- Ranges from 20 years old to 80 years old
- Average age is 49 years old
- Discretized age into 6 age groups (20s, 30s, 40s, 50s, 60s,70s)
- Visually, age does not seem to change brand preference
- Customers in 30s and 50s may have a weak preference for Acer

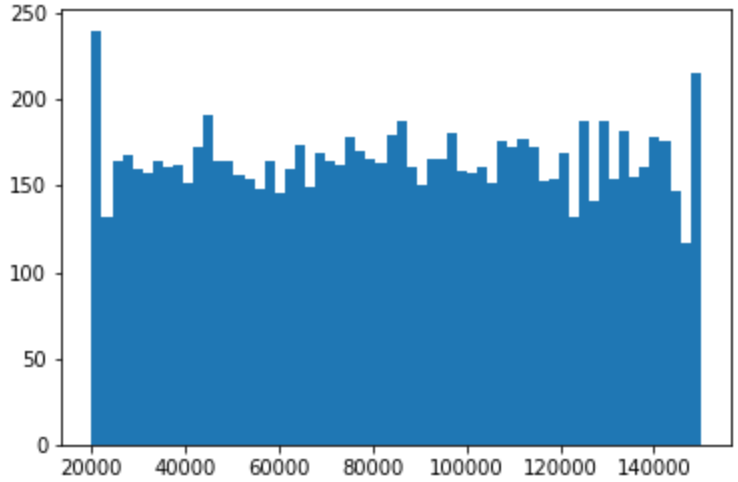
- Customers in their 70s may have a weak preference for Sony

Figure 3: Kernal Density Plot of Customers' Ages by Brand Preference



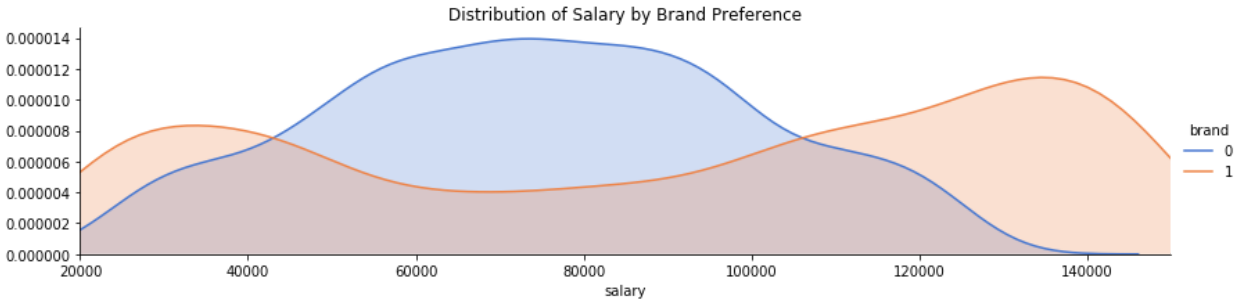
Salary

Figure 4: Histogram of Salary



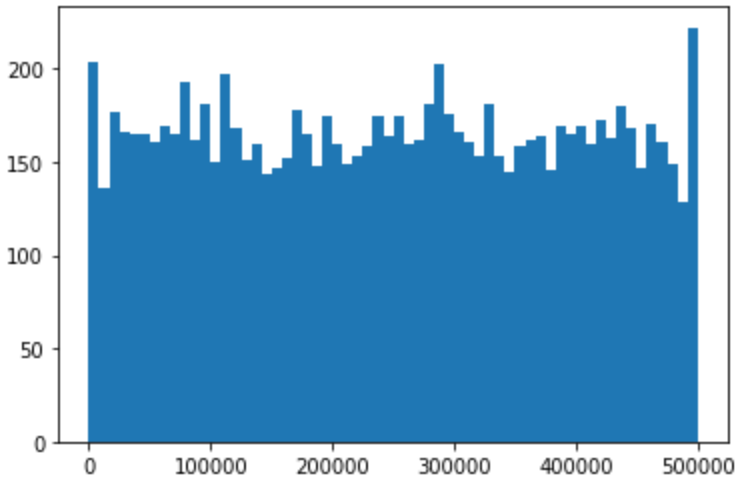
- Ranges from $20,000 to $150,000
- Average customer salary is $84,870
- Customers making between $50,000 and $110,000 may strongly prefer Acer
- Customers making less than $45,000 and more than $110,000 may prefer Sony

Figure 5: Kernel Density Plot of Customers' Salaries by Brand Preference



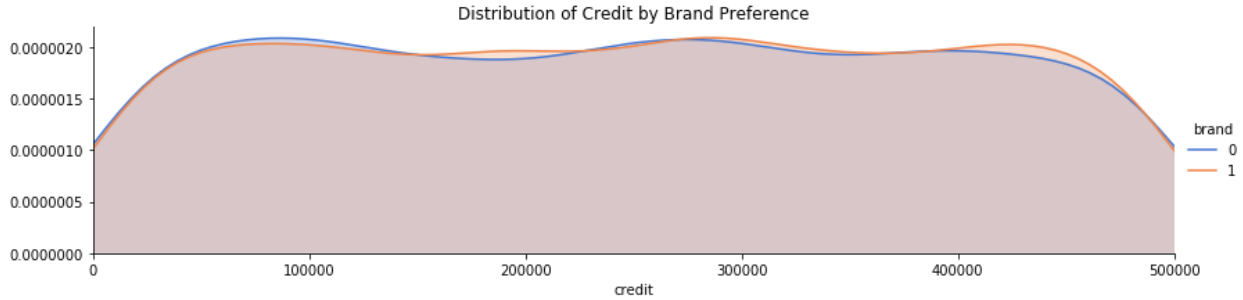
Credit Available

Figure 6: Histogram of Credit



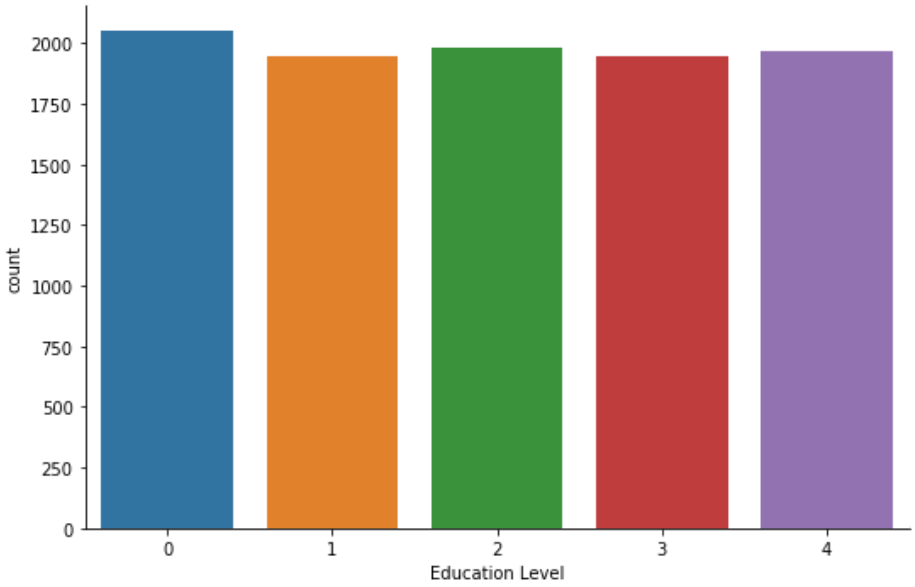
- Ranges from 0 credit $500,000
- Average credit available is $249,175
- No obvious relationship between credit and brand preference

Figure 7: Kernel Density Plot of Customers' Credit by Brand Preference



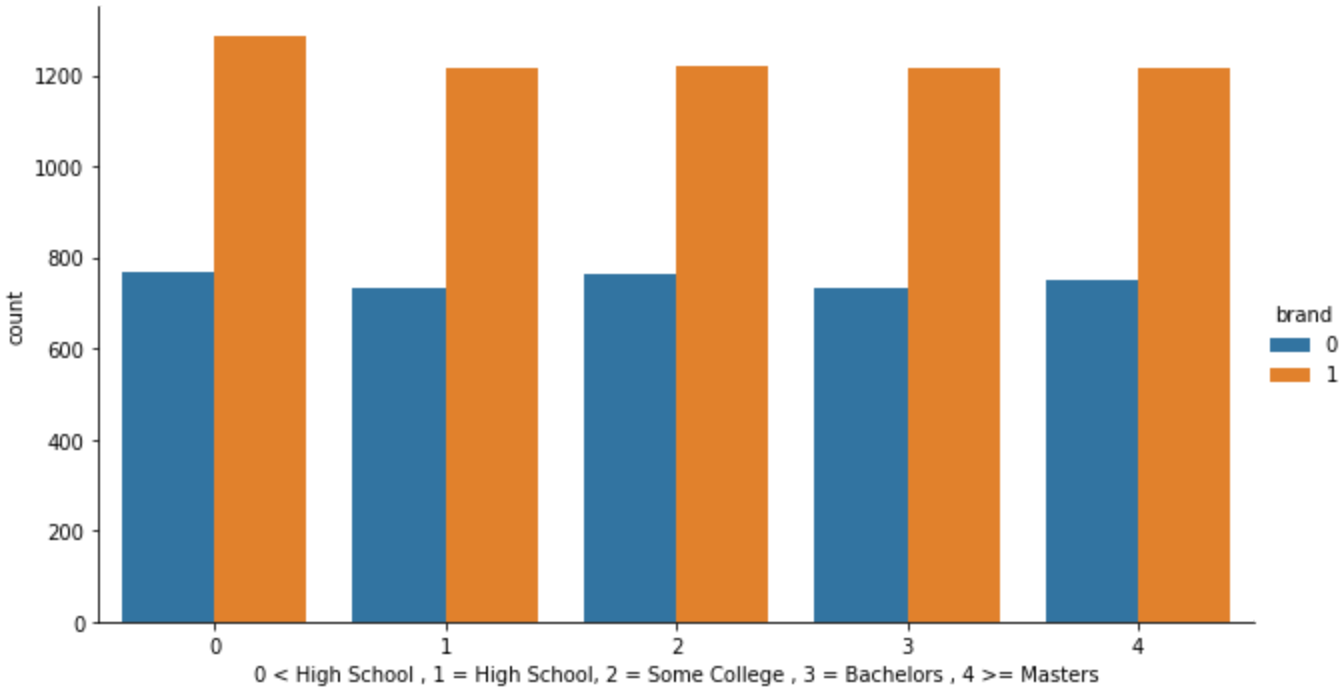
Education Level

Figure 8: Barchart of Education Levels



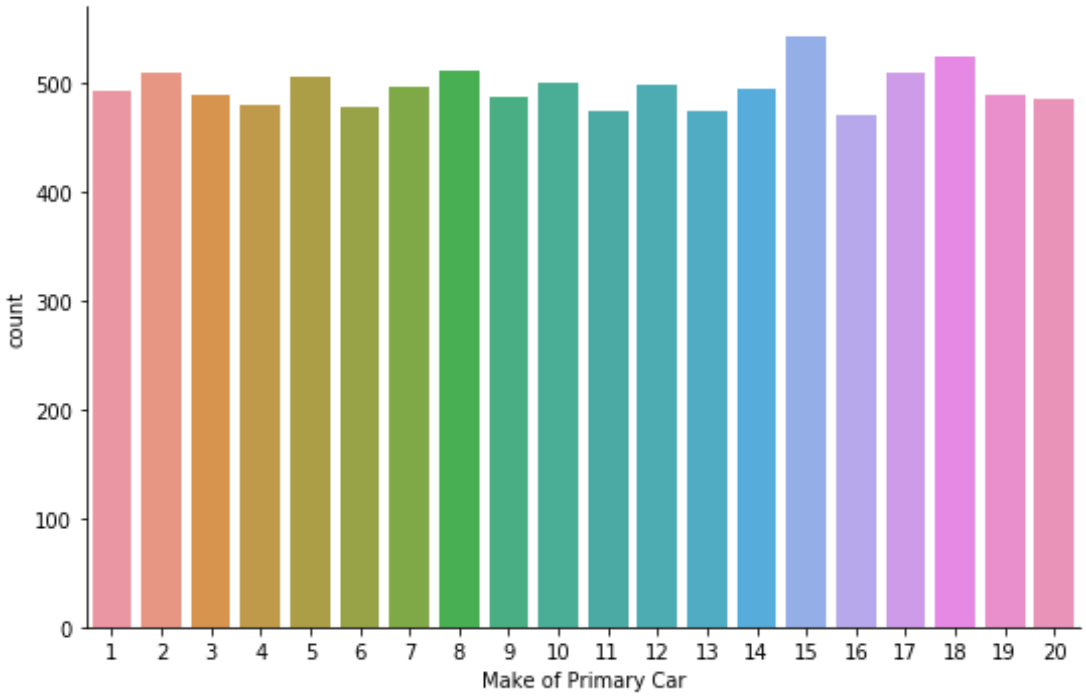
- 0 = Less than High School Degree
- 1 = High School Degree
- 2 = Some College
- 3 = 4-Year College Degree
- 4 = Master's, Doctoral or Professional Degree
- No obvious relationship between education and brand preference

Figure 9: Barchart of Education Levels and Brand Preference



Car Make

Figure 10: Barchart of Car Make



Key	Car Make
1	BMW
2	Buick
3	Cadillac
4	Chevrolet
5	Chrysler
6	Dodge
7	Ford
8	Honda
9	Hyundai
10	Jeep
11	Kia
12	Lincoln
13	Mazda
14	Mercedes Benz
15	Mitsubishi
16	Nissan
17	Ram
18	Subaru
19	Toyota
20	Other

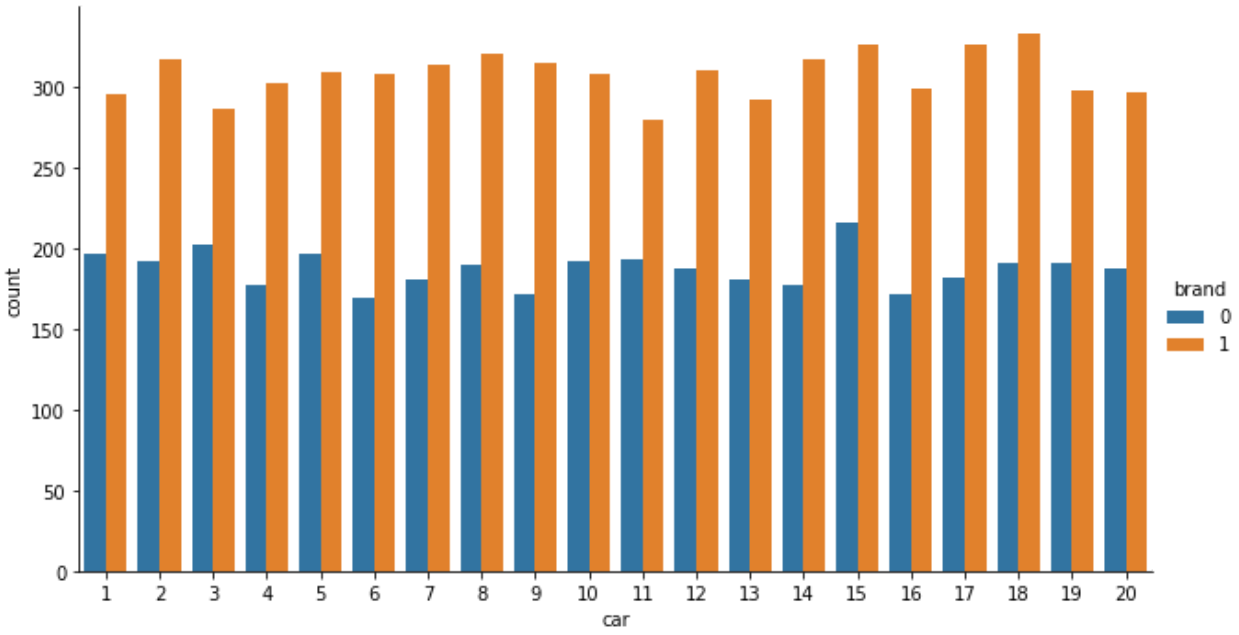
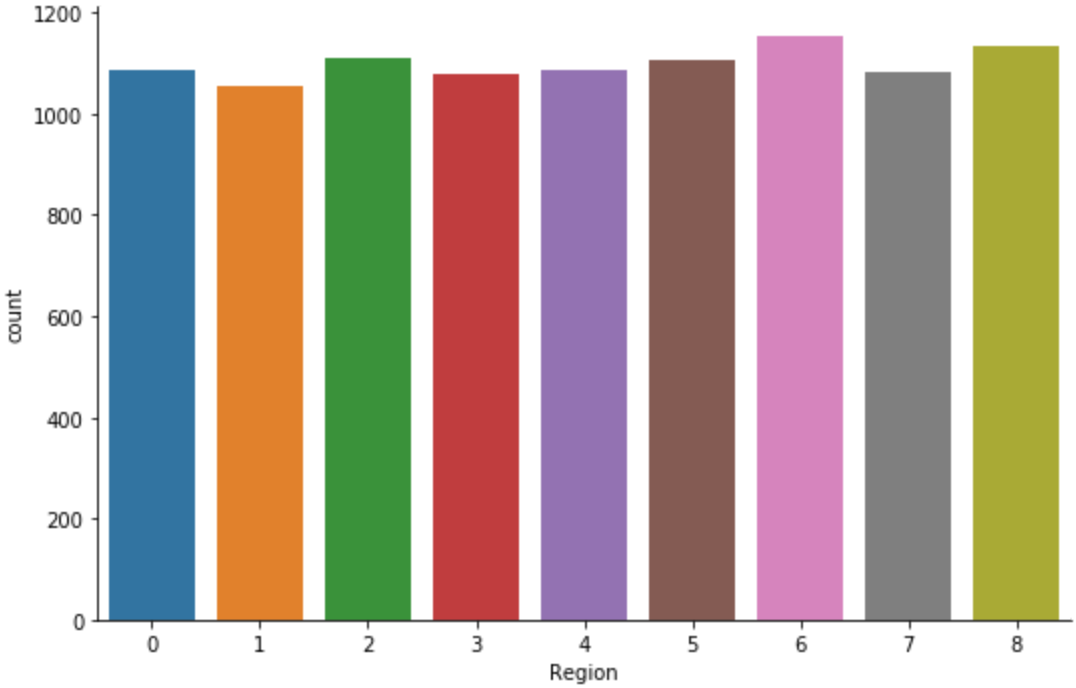


Figure 11: Barchart of Car Make and Brand Preference

- Make of customer's primary car
- No obvious relationship between car make and brand preference
- More customers drive Mistubishi cars than any other make

Region

Figure 12: Barchart of Region

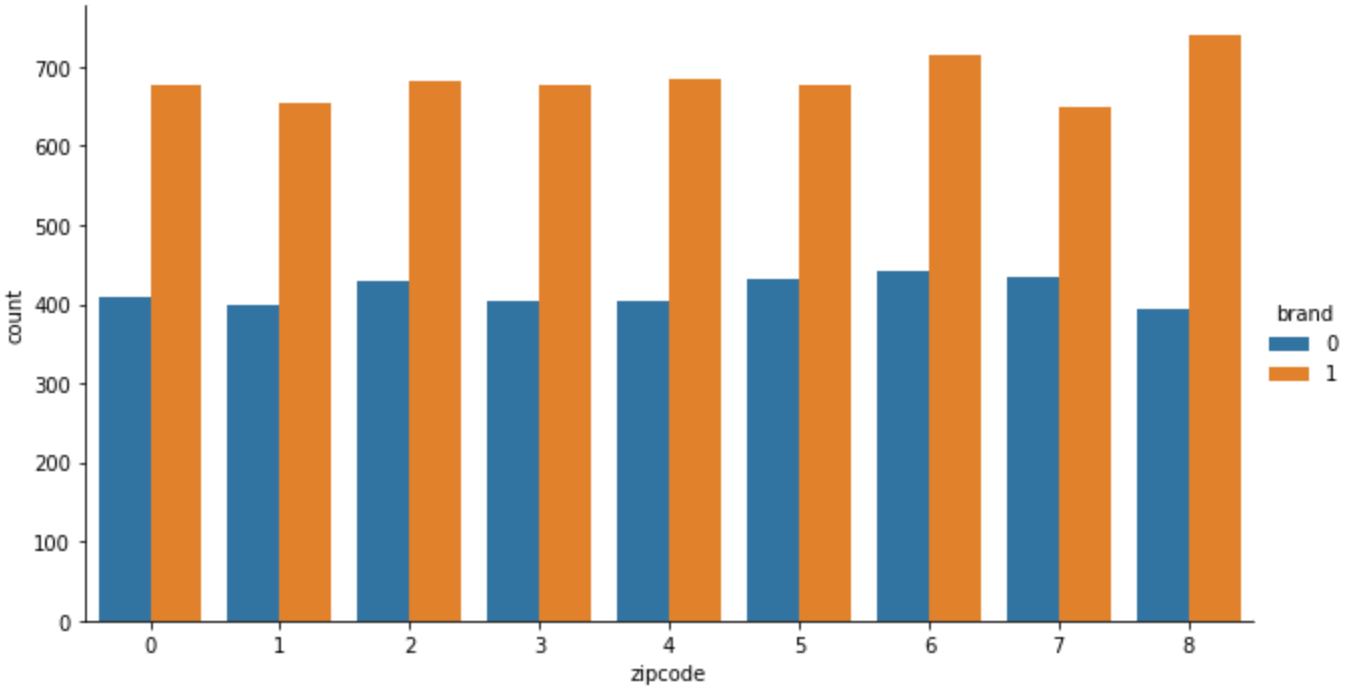


- 8 Regions derived from customer zip codes
- No obvious relationship between region and brand preference

Key	0	1	2	3	4	5	6	7	8

Region	New England	Mid-Atlantic	East North Central	West North Central	South Atlantic	East South Central	West South Central	Mountain	Pacific

Figure 13: Barchart of Region and Brand Preference



3. Results

Overall, our customers prefer Sony computers. We built a model that can **predict customer brand preference correctly 92% of the time**. The most influential variable was salary, which makes sense considering the variance in the plot of salary and brand preference (see Figure 5), compared to the other bivariate plots.

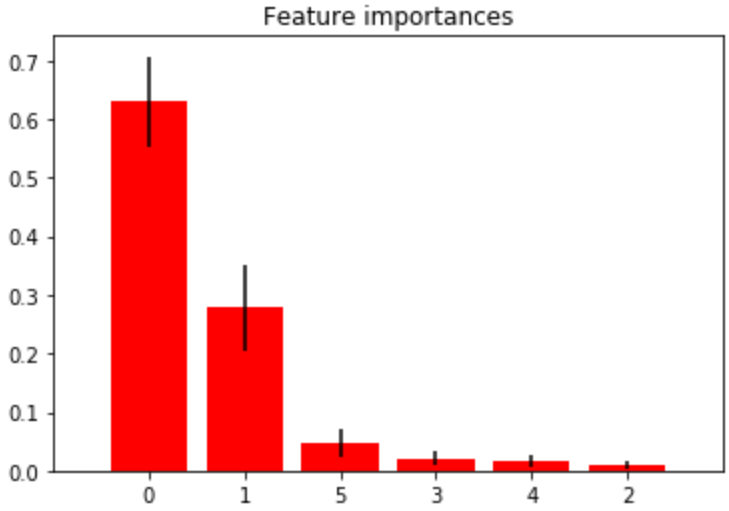


Figure 14: Barchart of Most Important Variables

Feature ranking:

1. Salary
2. Age
3. Credit
4. Car
5. Region
6. Education Level

4. Model Performance Metrics

Model	Binning	Class Imbalance	Tuning	Cross-Validation	F1	Accuracy	Kappa	Precision	Recall
Support Vector Machine					0.75971	0.61253	0	0.61253	1
DecisionTree (CART)					0.9162	0.89778	0.78518	0.92016	0.91227
Random Forest		UnderSampling			0.92825	0.91556	0.82607	0.96779	0.89182
Random Forest	Age				0.93355	0.91879	0.82914	0.93572	0.9314
Random Forest			Yes	KFold	0.93741	0.92251	0.83555	0.94042	0.93448
Random Forest		OverSampling			0.9375	0.92485	0.84336	0.95548	0.92018
Random Forest	Salary				0.93754	0.92404	0.84065	0.94444	0.93074
Random Forest			Yes	Stratified	0.93785	0.92311	0.83707	0.94261	0.93321
Random Forest					0.93895	0.92566	0.84392	0.94459	0.93338
Random Forest			Yes		0.94156	0.92848	0.84943	0.9425	0.94063

Figure 15: Confusion Matrix for Random Forest with Stratified 10-Fold Sampling