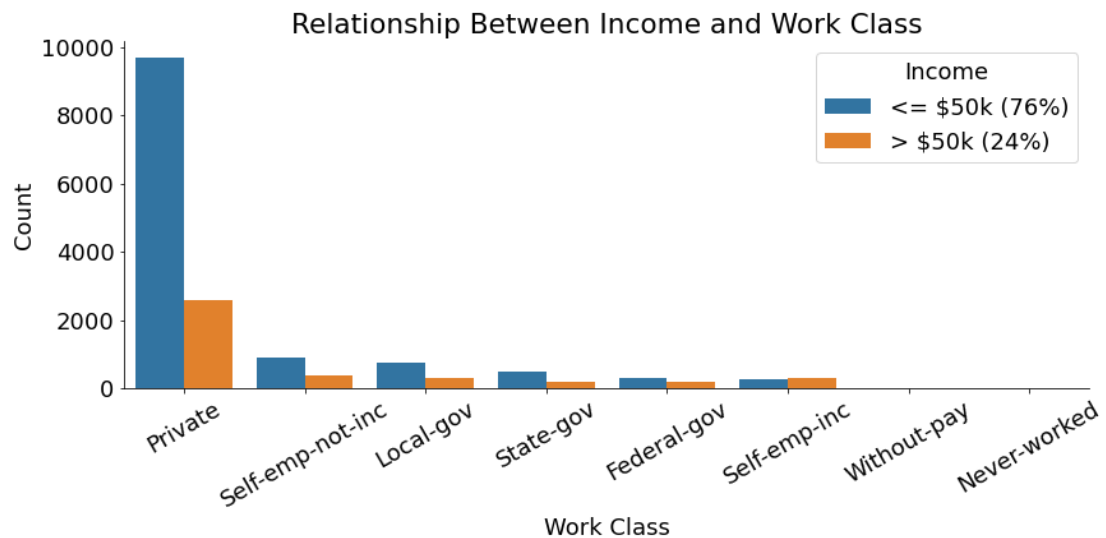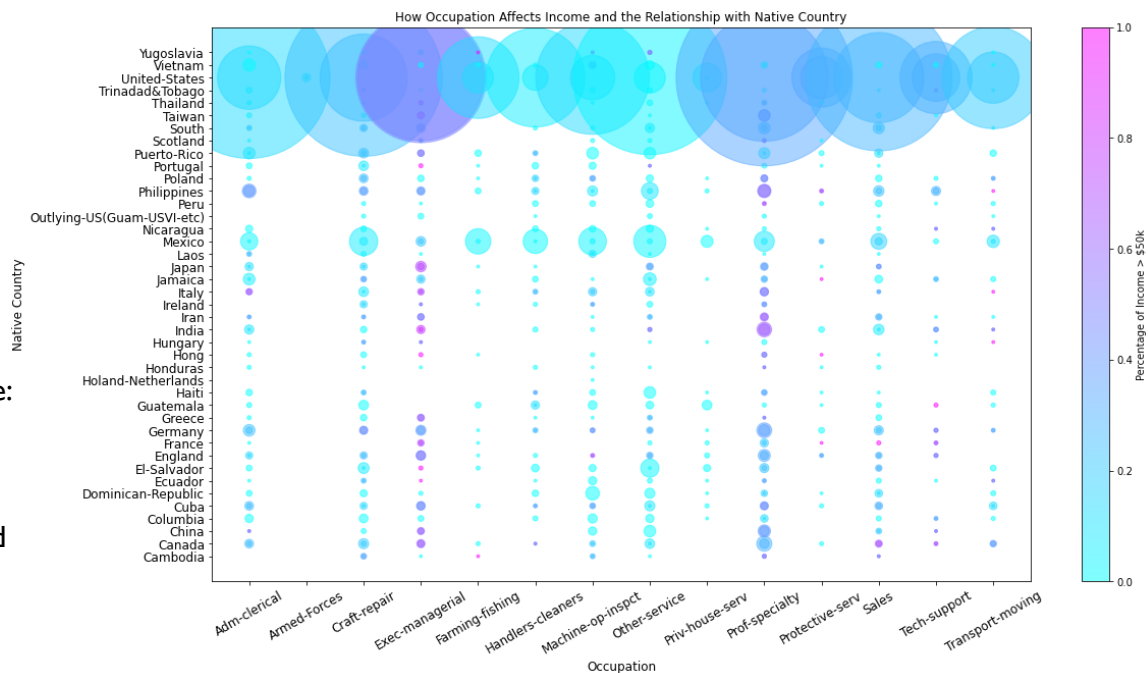| Question	As a member of the marketing team at UVW, I would like to know the relationship between income and workclass

- Regardless of income level,
  - the majority of people work in private companies
- Except for self-emp-inc,
  - the rest of the work classes has more people earning $\leq \$50k$ than $> \$50k$

Relationship Between Income and Work Class

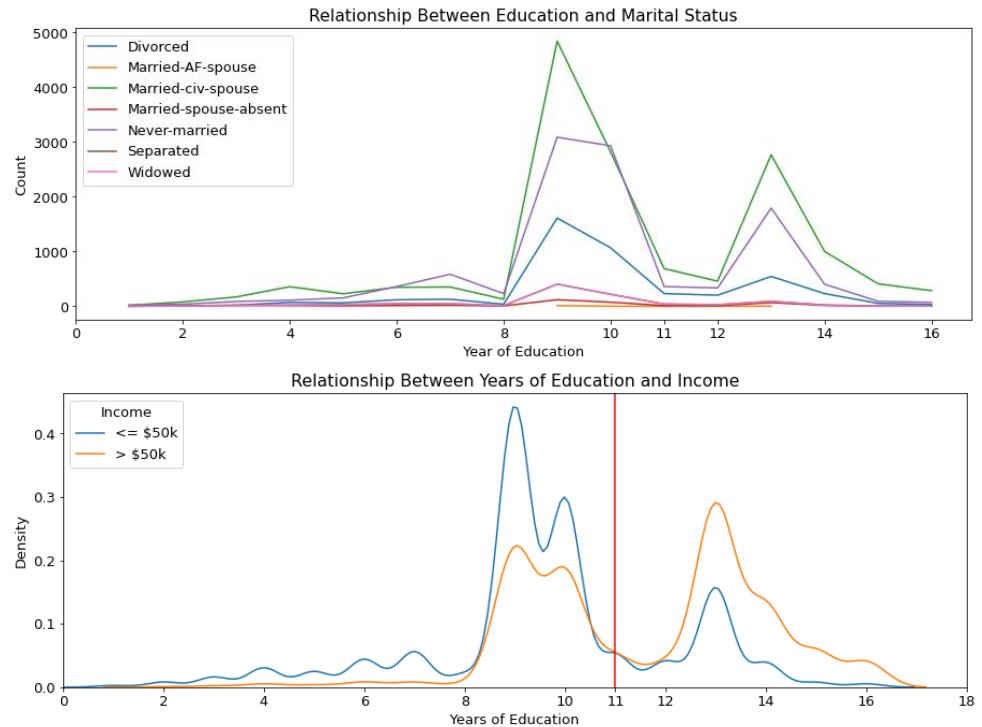| Question	I would like to know if occupation affects income and the distribution of native country

- There's a bias in this dataset
  - 91% U.S.-native
- The pinker the color, the higher the % of high income
- People with high income feature:
  - Non U.S-native,
  - Executive management, and
  - Prof-specialty

How Occupation Affects Income and the Relationship with Native Country

The worker wants to know the relationship between education and marital status
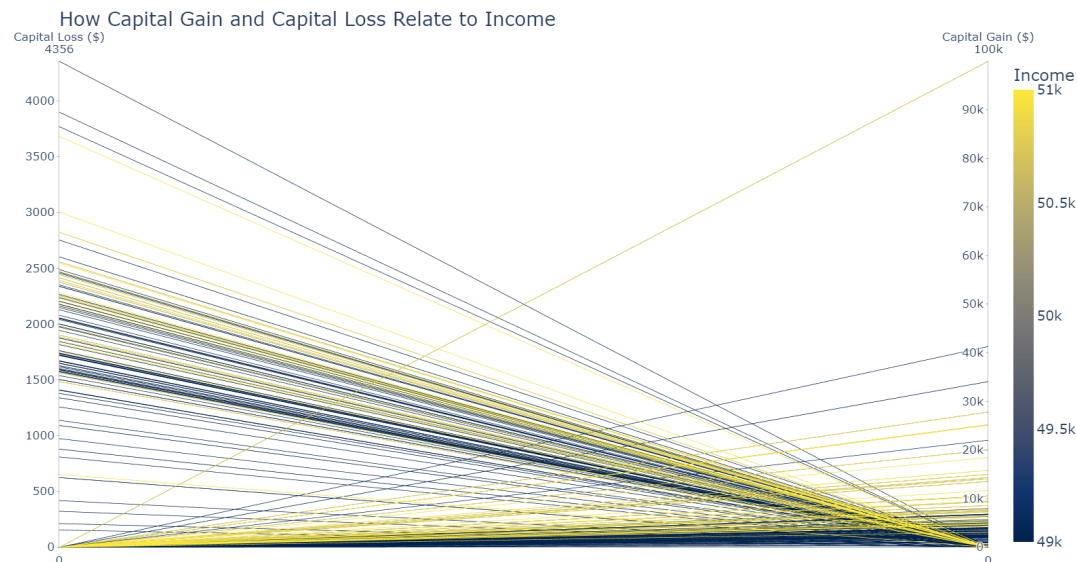
- Regardless of marital status,

  o most of them have 9~10 years or

    13 years of education

- 11-year education is a watershed

  o Under 11 years, more likely to

    earn less than $50k and vice sersa

Relationship Between Education and Marital Status

Relationship Between Years of Education and Income

The Director of Marketing would like to know how capital gain and capital loss relate to income

- Yellow line → High income

  Blue line → Low income

- People who have capital loss and

  capital gain have no overlap

- The amount of capital gain is much

  more than the amount of capital loss

How Capital Gain and Capital Loss Relate to Income

Analysis of Different Income Levels

76% of people earned less than $50k

- Most of the people worked in private companies

  Regardless of high or low income → **Workclass**

- Most people work 40 hours a week
- More than **45 hours** → more likely to earn more than $50k → **Hours-Per-Week**

- People with 9, 10 or 13 years of education are the mainstream
- More than **11 years** → more likely to be in higher income groups → **Education-Num**

- The **age of 35** is the watershed
- Low-income earners → Between the ages of 15-35
- High-income earners → Between the ages of 35-60 → **Age**

- People with higher capital gain or loss amounts are high-income earners. → **Capital-Gain Capital-Loss**

Which Factors Should Be Used?

Use Application to Predict High Income

- Identify whether people have capital transactions

  No ⟶ / Yes ⟶

  No
  - ✓ **Marital Status**
  - ✓ **Age**
  - ✓ **Education-Num**
  - ✓ **Relationship**

  Yes
  - ✓ **Marital Status**
  - ✓ **Age**
  - ✓ **Education-Num**
  - ✓ **Capital-Gain**
  - ✓ **Capital-Loss**

Manually Cluster High Income People

- ✓ **Hours-Per-Week**
- ✓ **Age**
- ✓ **Education-Num**
- ✓ **Capital-Gain**
- ✓ **Capital-Loss**

Not Recommend to Use

- × Race

  85% white

- × Native-country

  91% U.S.-native

  } Data is bias

- × Fnlwgt

  No significant differences among people of different income levels

Final Weight and Income Level

Density — Final Weight (1e6)

Income: > $50K, <= $50K