CDSS Hackathon
Differential Privacy

Andrea, Lucy, Mori, & Sunny

Mission statements

Measure service Quality

Aggregate average of tip to fare ratio from each transaction reflects the service quality of each vendor

User privacy

Avoid association between identity and tip amount through **anonymization**

Maintain accuracy

Minimize the difference between real and private value with low epsilon value

Visualization

Present the data through bar chart

Problem vs solution

Problem

The Cab company wants to measure the **service quality** of the two vendors. We want to preserve **customer privacy** while extracting data.

Solution

Tip-to-fare ratio serves as a **proxy** to measure cab ride service quality. Generate **synthetic data** for tip and fare amount using **differential privacy algorithm (2-margin).**

Challenge

The trade off between accuracy and privacy

Epsilon value measures the privacy level as increase in epsilon decreases privacy.

As we add more noise to the data, the accuracy decrease and privacy increase.

We want to increase the privacy while making sure the synthetic data still reflect attributes of the population

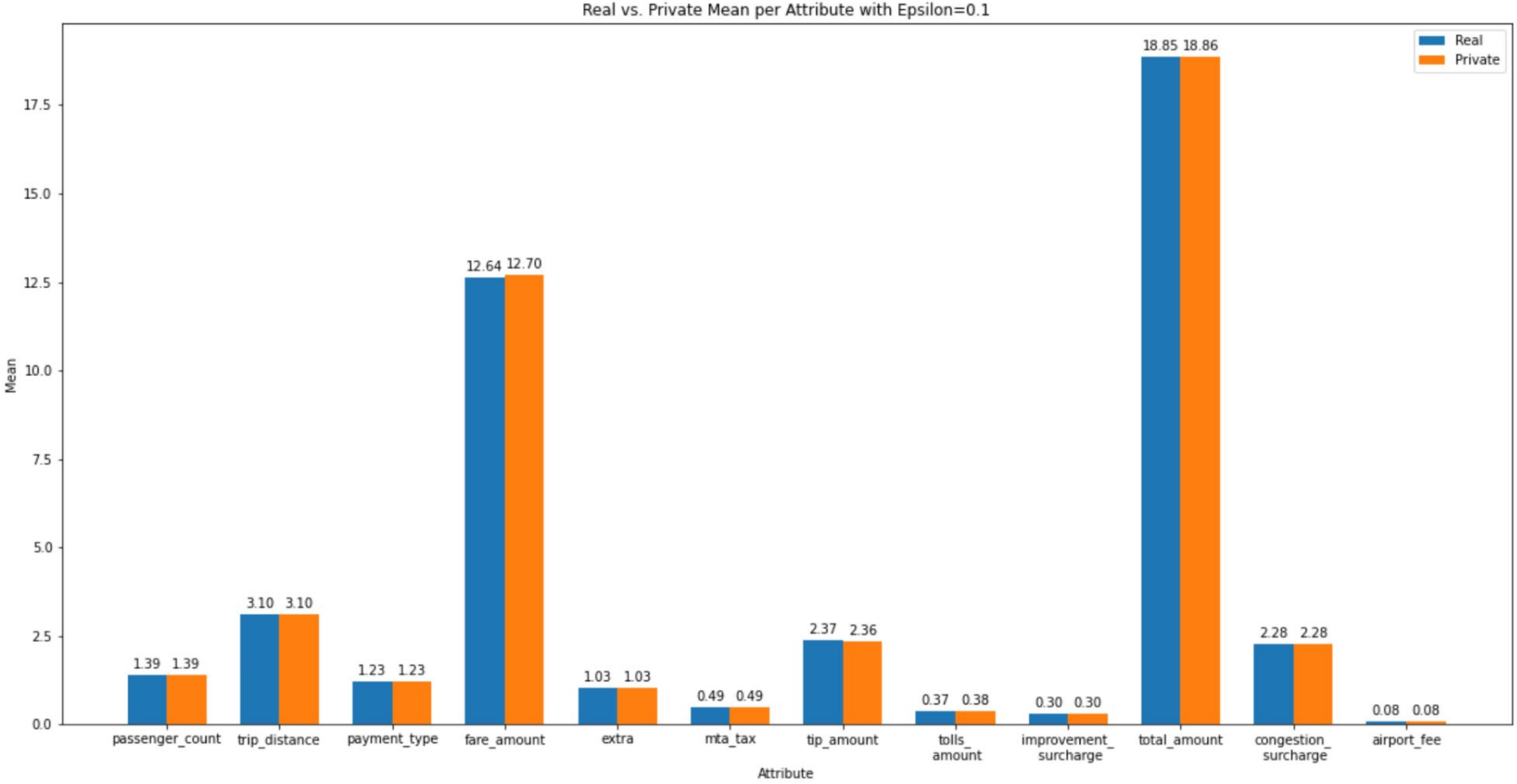
Methodology

- Use the PyDP library to measure statistics after introducing noise to our data, which creates synthetic data that preserves the descriptive statistics of the original data while obscuring individual's information
- Calculate the mean tip amount for each cab ride as a percent of the fare amount
- Adding **noise** to the tip-to-fare ratio
- Compare mean tip percent between vendors to compare quality of service using t-test
- Compare t-test results before and after differential privacy to test accuracy

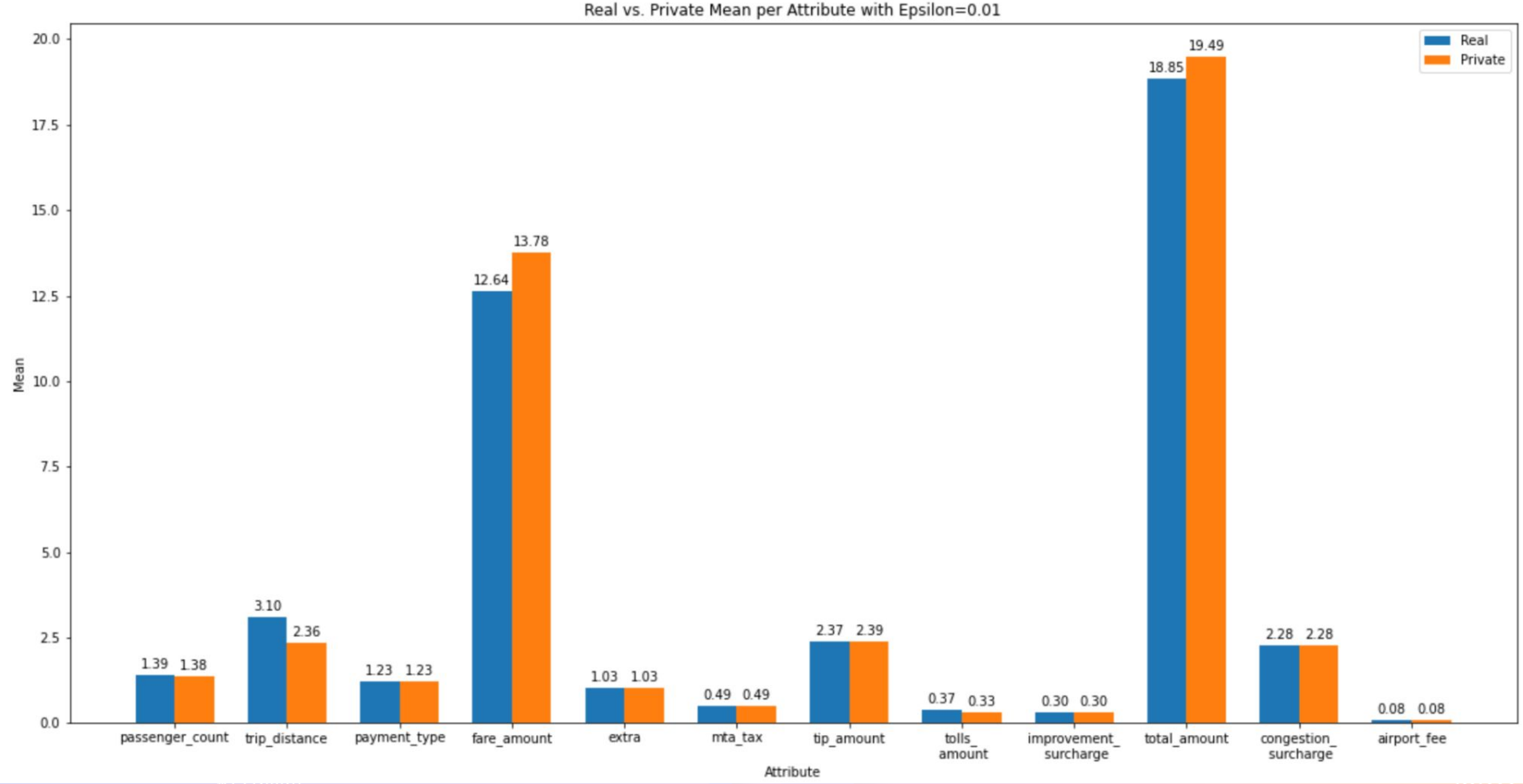
Methodology

- Why PyDP?
  - Generates synthetic dataset without too much dependence on the underlying dataset (i.e., only end-results of aggregated information are produced)
  - It utilizes Laplace mechanism and has built-in functions on creating bounded means,

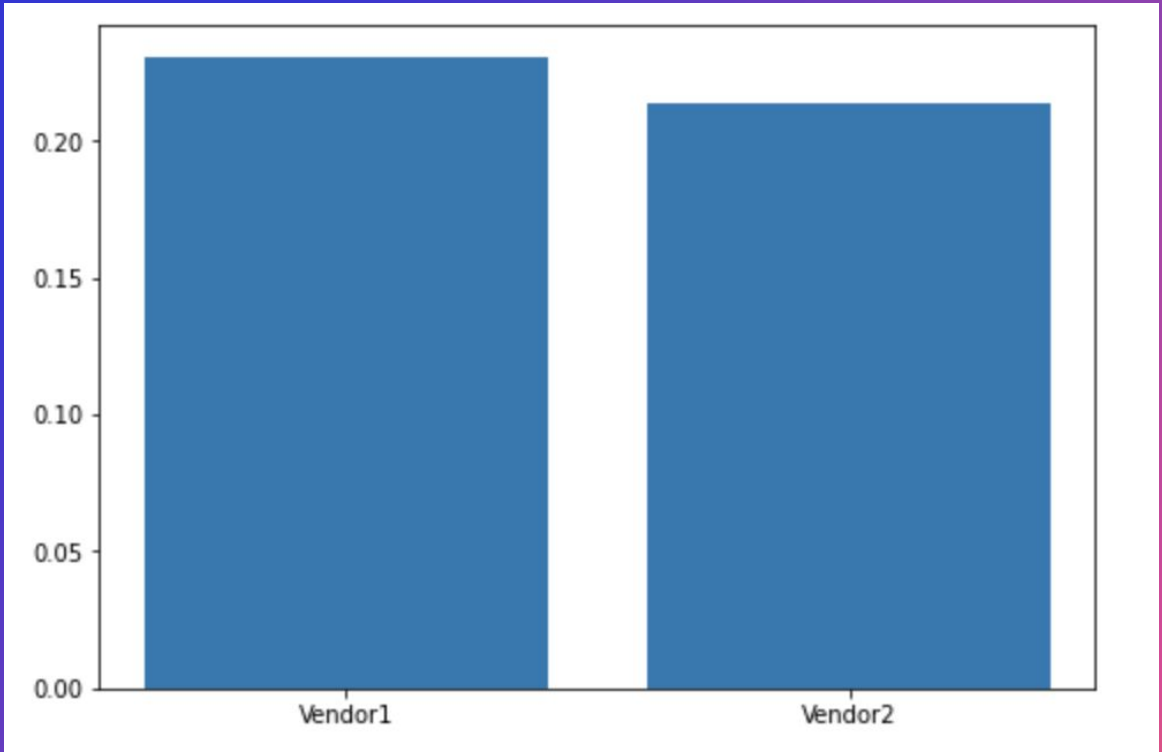
Real vs. Private Mean per Attribute (ε = 0.1)



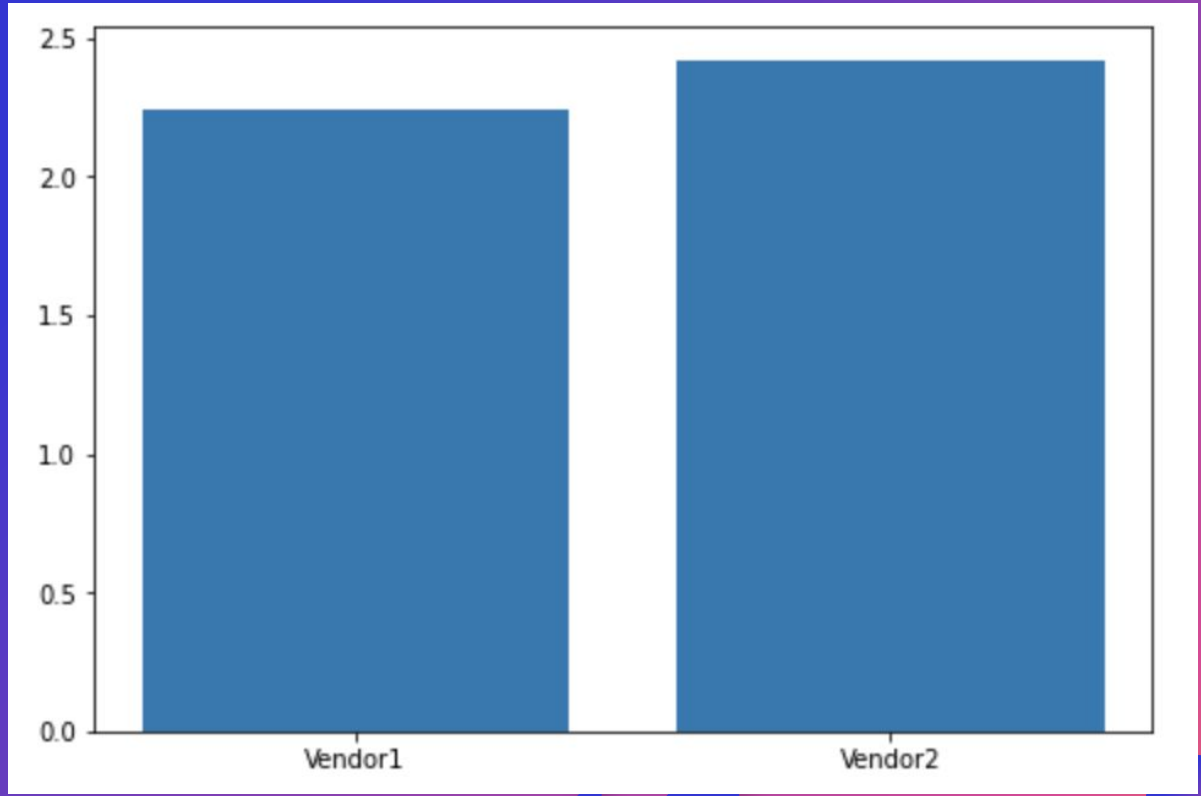
Real vs. Private Mean per Attribute (ε = 0.01)

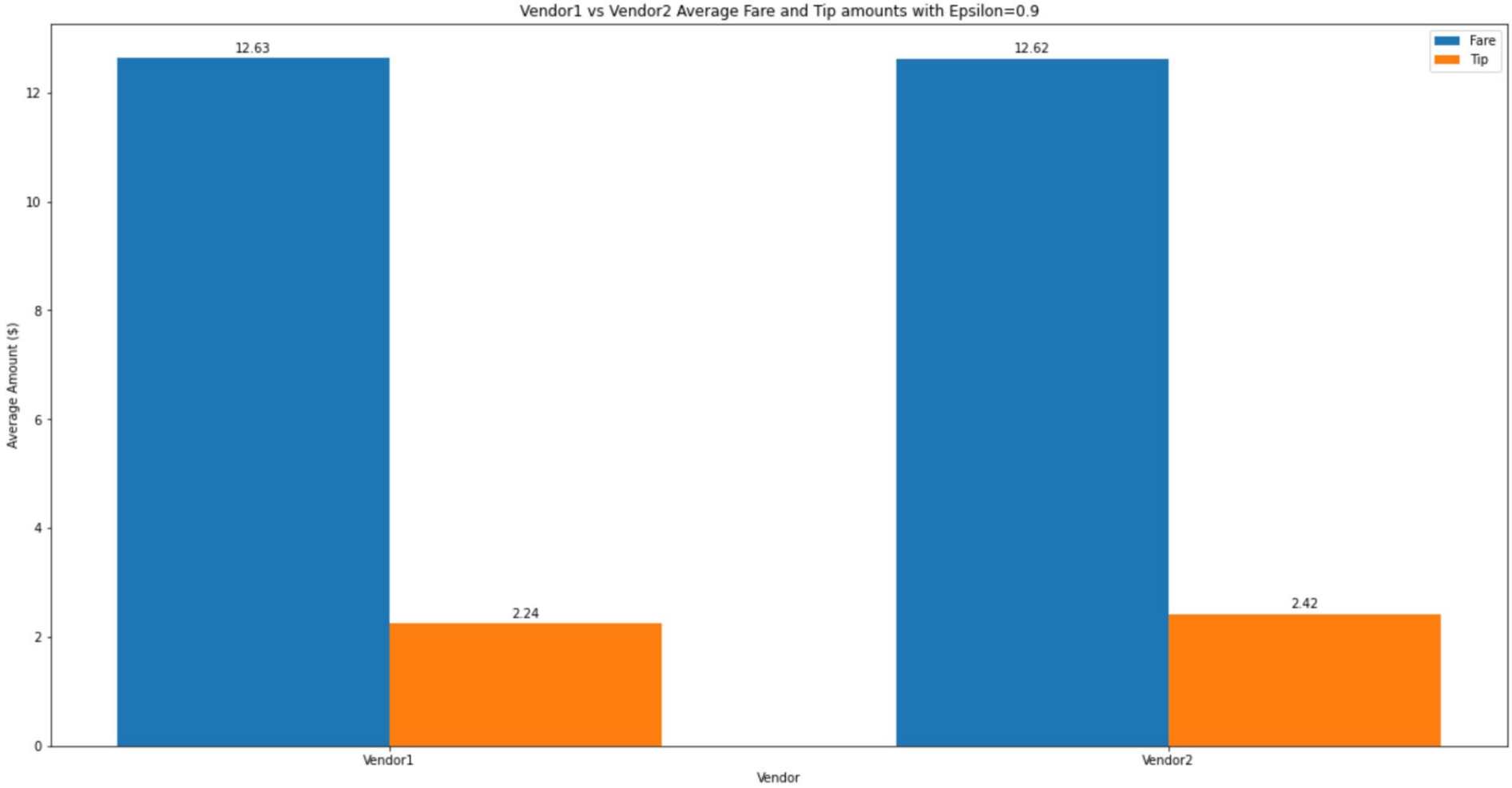


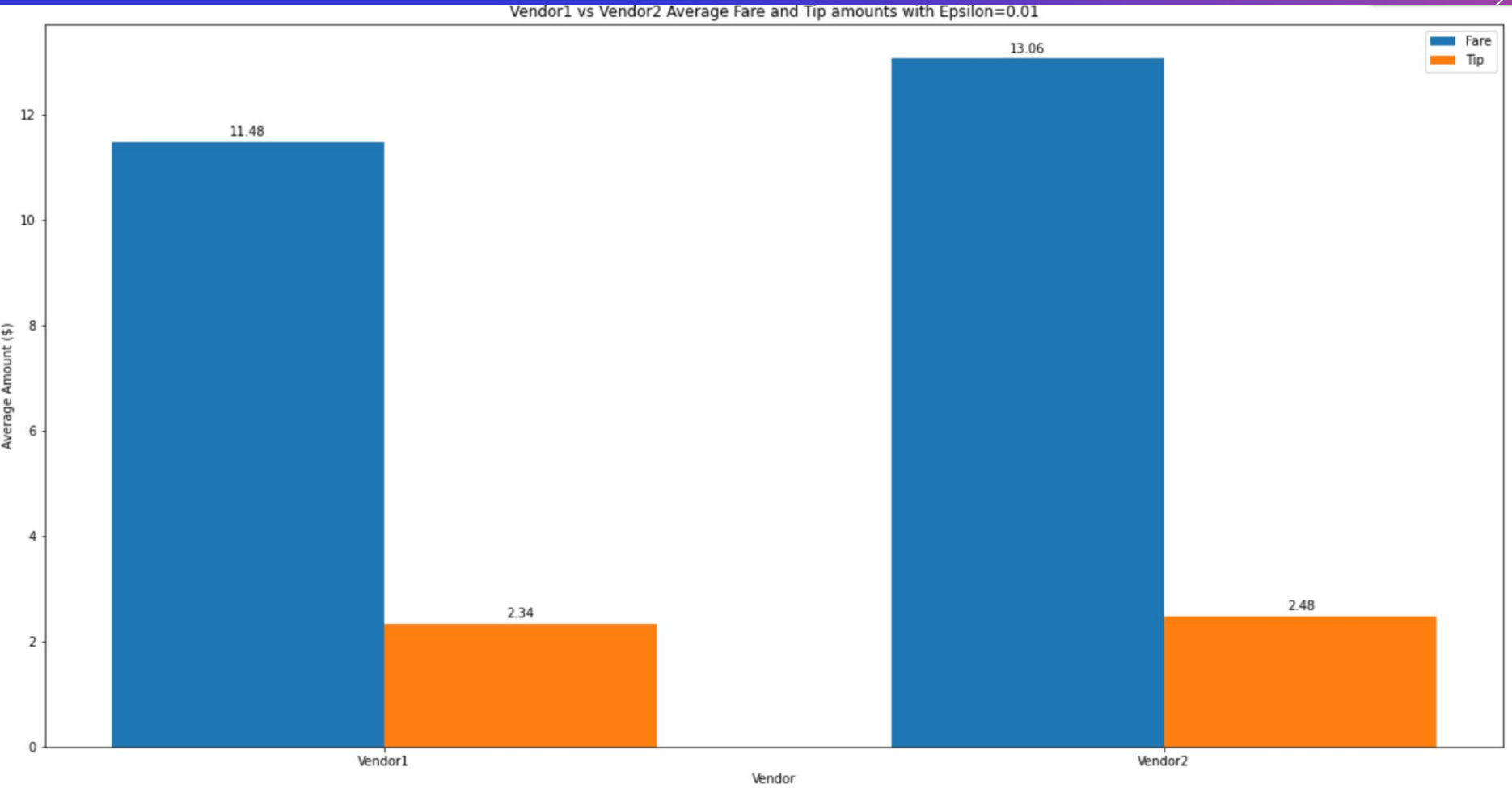
Average tip amounts
Vendor 1 vs Vendor 2 (ε = 0.9)



Average tip ($) amounts
Vendor 1 vs Vendor 2 (ε = 0.9)




Vendor1 vs Vendor2 Average Fare and Tip amounts with Epsilon=0.9
Fare
Tip
12.63
2.24
12.62
2.42
Average Amount ($)
0
2
4
6
8
10
12
Vendor1
Vendor2
Vendor


Vendor1 vs Vendor2 Average Fare and Tip amounts with Epsilon=0.01
Fare
Tip
11.48
2.34
13.06
2.48
Average Amount ($)
0
2
4
6
8
10
12
Vendor1
Vendor2
Vendor