

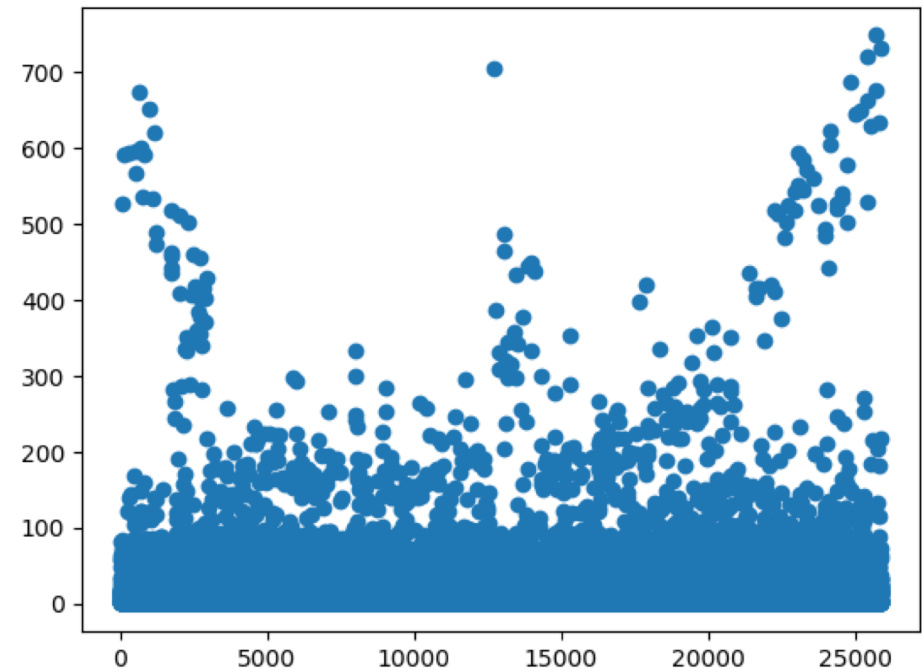
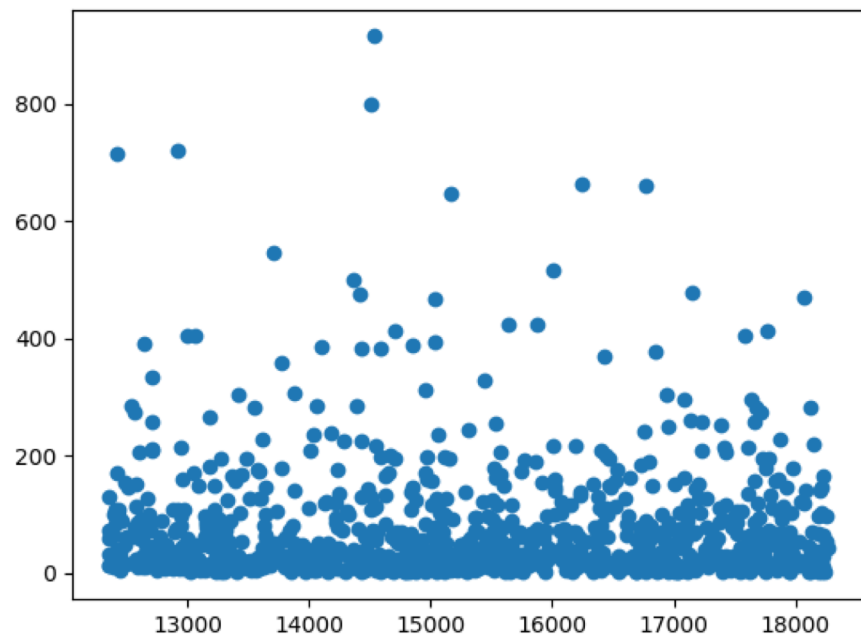
夏課題

千葉 泰治

夏課題内容

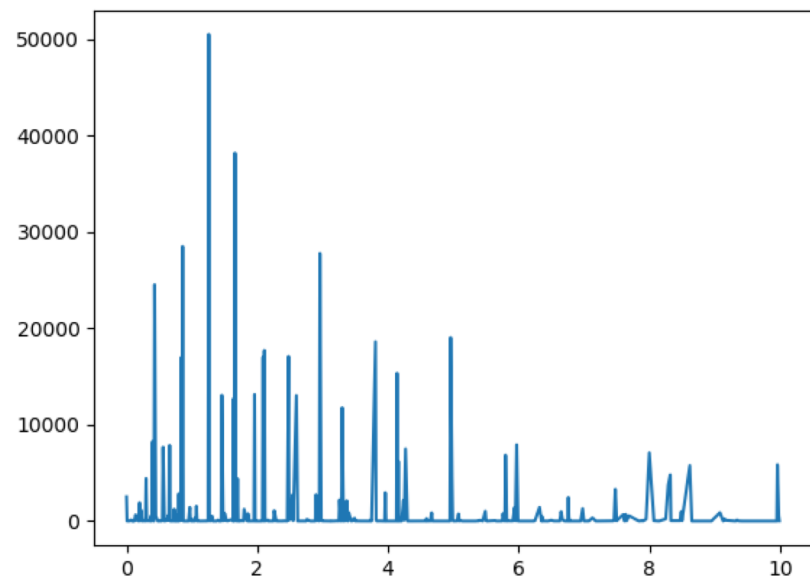
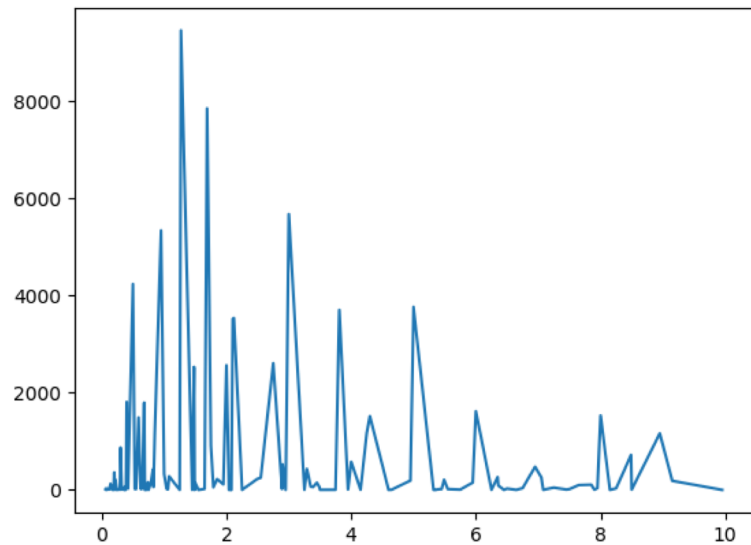
- 先年度のPWSCup(匿名加工技術の腕を競い合う学生向けの大会) で利用された購買データと今年度自分が扱う予定のデータについての比較を行った。

ユーザーごとの購買回数



- 2つとも基本的に購買回数に偏りがあるが、このグラフで表示していない圧倒的な外れ値が左図 (Vshopper)のほうが多かった。
- またその分値の多様性は右図(Online retail)のほうが見られた。

価格帯ごとの比較



- 双方ともに2ドル以下の商品に偏りがあった。
- こちらも先ほど同様に右図のほうが値の範囲に広がりがあった。

終わりに

- グラフの種類をフィーリングで選んでしまったが、特に2つ目は明らかに間違った手法を選ってしまった。
- いくつかの他のグラフでも試したが、すべて微妙だったので、このように値が飛び飛びになるデータの比較方法についても今後は検討したい。

- またレコード数の差も大きいというえにVshopperデータのほうに圧倒的な外れ値がいくつもあり、データの数に偏りがあったためグラフで見た時にOnline retailデータのほうが多様性があるように見えるのも問題だった。