

## 一日目, Sep.1 火曜日

インターンシップ説明書を参加する

解析例とサンプルコードを読みます

## 二日目, Sep.2 水曜日

データを読み込む

データはtxtタイプです、まずglobというlibraryを使って、フォルダの全部のtxtのファイル名を取って、pandas libraryのread\_csvとconcat関数でデータを読み込む

## 三日目, Sep.3 木曜日

データの評価:

サイズ 563883行 8列、

タイプ 欠損値: なし

重複した値: 4 9 9 行

## 四日目, Sep.4 金曜日

データのcleaning: - 重複した値 - date特徴 - time特徴

新しい変数を作る - month, day, hourJP - country, countrycode, ipcountry - region, city - latitude, longitude - timezone, hourLocal - method, http\_version

## 五日目, Sep.7 月曜日

各特徴について、探索的データ分析を行います、可視化の手段でパターンを認識し、正式な分析のために準備します

## 六日目, Sep.8 火曜日

中間報告会に参加する

今までやった内容を整理する

## 七日目, Sep.9 水曜日

推論統計 - 月について chi square test

honeypotに関する論文を探す

## 八日目, Sep.10 木曜日

正式な分析を行います

- 時間に関する特徴
- 場所に関する特徴
- テクニックに関する特徴

## 九日目, Sep.11 金曜日

正式な分析を行います

- 時間に関する特徴
- 場所に関する特徴
- テクニックに関する特徴

## 十日目, Sep.14 月曜日

最終の報告を準備する

最終報告会

## Sep.15 - Sep.28

分析の修正、改善

インターンシップについての感想