グラフ, 散布図

自動的に生成された説明

ボランティアは上下がかなり激しい　→　休日に多くなるのでは？と思ったが下記の図から休日に特に多くなるわけではないとわかった。　→　「疑問」ではなぜ上下が激しいのか？　特に10月15日、16日だけで1000件近くのデータが取得されている。→　なにかイベントごとがあったのかな？12/15現在　進捗無し

公式から引用　「ボランティアのマッパーからなる少人数のクルーは、冬にも調査を行った。」　動員数が少ないので2015年の冬以降は件数が少ないのである

グラフ, 棒グラフ

自動的に生成された説明

TreeCountStaffとNYC parksStaffについては平日に比べて休日のほうが少しデータ取得件数が少ないといえる。どのタイプにおいても水曜日が一番件数が多いが10.15が水曜でこの日だけで400件ほどデータの取得があったのでそれのせいだと思われる。

余談

NYCの公式を見たがvolunteerのデータはstaff同様に正確だと書かれていた

グラフ, ヒストグラム

自動的に生成された説明グラフ, ヒストグラム

自動的に生成された説明

グラフ, ヒストグラム

自動的に生成された説明

これのみではhealthは判断することができない

Spc\_commonごとに分布は違うであろう　木の胸高直径をはかっているのであるならば特徴として種的に太い、細いがあると思うからだ

木に近縁種かなり混じっている　例　white oak black oakすべてオーク木の仲間らしい

もしかたら近縁ごとまとめたものを特徴量として用いてもいいかもしれ

グラフ, ヒストグラム

自動的に生成された説明グラフ, ヒストグラム

自動的に生成された説明

グラフ, ヒストグラム

自動的に生成された説明

分布はあまり変わらず３つを分けるのは難しそう

グラフ, ヒストグラム

自動的に生成された説明

この図からhelth=0はtree\_dbhが低い値に集中しているので育ちのわるい木なのではないか？

公式の資料からまずまず基準を探ってみる

公式から引用　「樹木の健康状態に関する利用者の認識を示す。」　利用者の主観によるものである可能性があるので明確な基準がないのかな？

人の主観によるものならばその日の天気によって感じる雰囲気が異なると思われるので天気が重要？

葉をつけ、冬は枯れる木であるなら元気なのに枯れてしまっている木については、状態が悪くつけられてるかも

Steward この木に観察されたスチュワードシップのユニークな兆候の数を示す。切り株や枯れ木には記録されない。

Problemsについてわかったこと

Guards ガードの有無と、使用者が役に立つガードだと感じたか、有害なガードだと感じたかを示す。枯れ木や切り株の場合は記録されない。 よってnull値は枯れ木、切り株かも！

Stones 樹木の根に敷石による問題があることを示す。

Metalgrade? 樹床に設置された金属製格子に起因する根の問題の存在を示す。

Trunk other他の根本的な問題の存在を示す

Trunk\_wire トランクに巻かれたワイヤーやロープによるトランクの問題の存在を示す。

BranchiLight 枝の中にある照明（通常はストリングライト）やワイヤーが原因で枝に問題があることを示す。→ クリスマスらへんは多いかも

**brch\_shoe**枝のスニーカーによる枝の問題の存在を示す。

なぜ欠損しているかのなぞが解けました

1,steward 14883

2,guards 14943

3,Problems 12243

1,2両方欠損 13007

すべて欠損 7323

切り株だと胸高直径はかれなくね？胸高直径少なくなるやろ

枯れ木であるならばその地域の雨量すくないんじゃね？　→　雨量を特徴量にするのでもどこからどこの期間を雨量とするんだい？

やく2/3の木が枯れているもしくは切り株の可能性

null

health

1 10368

0 2170

2 469 分布的にはどちらも変わらなそう

Notnull

1 5383

0 1365

グラフ, ヒストグラム

自動的に生成された説明グラフ, ヒストグラム

自動的に生成された説明2 229

わかることnot\_nullのデータはデータ総数のわりに5付近に集中している？　→　もしかしたら新しい木たちなのかな？

Both\_nullは寿命や病気で死んでる木が多いのでちゃんと育ってからしんどるんかな

Status追加

グラフ, ツリーマップ図

自動的に生成された説明

グラフ, ツリーマップ図

自動的に生成された説明

[0.3555, 0.3676, 0.3618, 0.3985, 0.3749] f1\_score 0.37166

0.00022114640000000033

[0.3551, 0.3673, 0.362, 0.3947, 0.3754] recall\_score\_all 0.3709

[0.3606, 0.3732, 0.3682, 0.4092, 0.3771] precision\_score\_all 0.37766

１と０のデータが判別しにくくなってしまった

1と２のデータのうまく分けられる特徴量だといえるね。特に１のデータを２と予測するもの現象がすごい

2で間違える中４０個ほど減らせるのであれば特徴量として用いていいいと思う

かといって2を１と間違えるわけではないので用いるべきだ2のrecallはだいぶ良くなっていると感じられる。

じゃーなぜ1と2が分けやすくなった特徴量なの？

おそらく1

てかもしかしたら

落葉樹は冬葉をつけないのでもしかしたらhelthが悪いの多いかと思ったがやmapleなどを出してみたがそんなことなさそう。下記の図はcherry

グラフ, ヒストグラム

自動的に生成された説明

木は潮風に弱いらしいうみに面してる面してないとかで分けるといいかもしれない