

Walmart Recruiting_Store Sales Forecasting

考察レポート

目次

1. 課題の概要
2. EDA
3. 評価指標
4. モデルの考察

1.課題の概要

walmartの店舗ごとの売上予測

- 店舗数:45店舗(3タイプ)
- 対象期間
 - トレーニング:2010-02-05~2012-10-26
 - テスト:2012-11-02~2013-07-26
- 商品は81のdeptが存在

Input

- 過去の売上
- 休日の情報
- 気温
- 燃料価格
- 消費者物価指数
- 失業率



Output

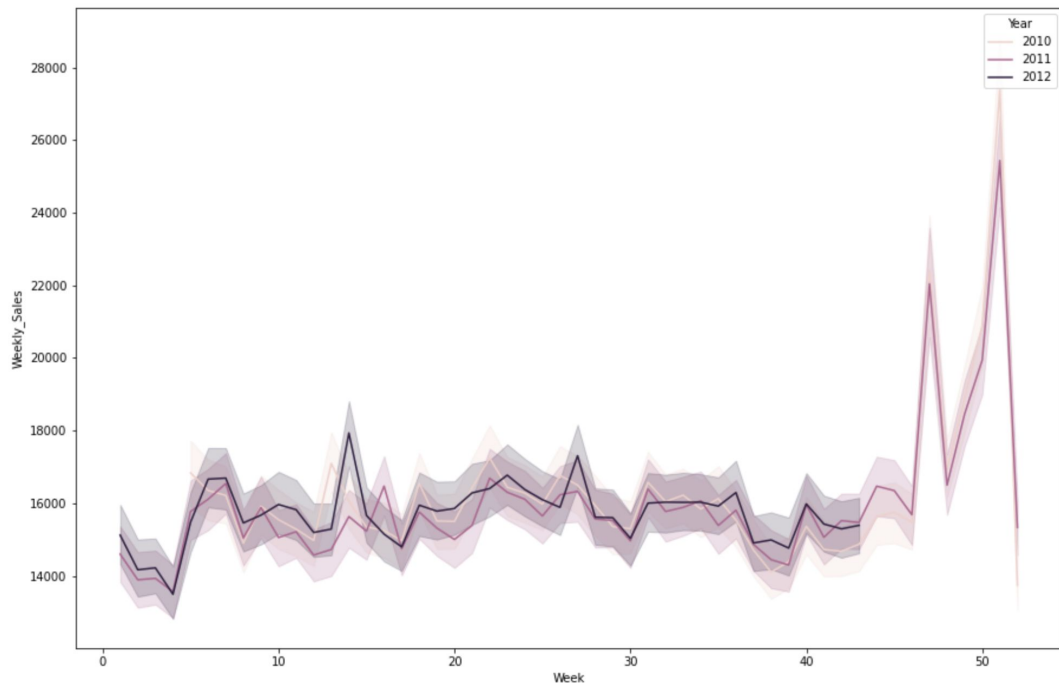
- テスト期間中の店舗、dept別の売上を予測

2.EDA(まとめ)

- 店舗は規模ごとにA、B、Cの3タイプに分類され、半数以上の店舗が1,500万以上の規模で、タイプAに分類される。
また、店舗の規模は売上に大きく影響しており、deptごとに売上の傾向が異なる。
- 特別休日の平均売上は平日より高く、2010年は2011年、2012年よりも売上が高くなっている。
- クリスマスは、1年の最後の週に該当しているが、最も高いのはその前の週の売上が高くなっている。そのため、特別休日はThanksgivingの売上が最も高くなっている。
- 1月は他の月に比べ、売上が最も低い。これは、Thanksgivingやクリスマスの影響がしており、消費者の購買行動が落ち着いたと考えられる。
- 燃料価格、気温、CPI、失業率は、週次売上高に影響を与えない。

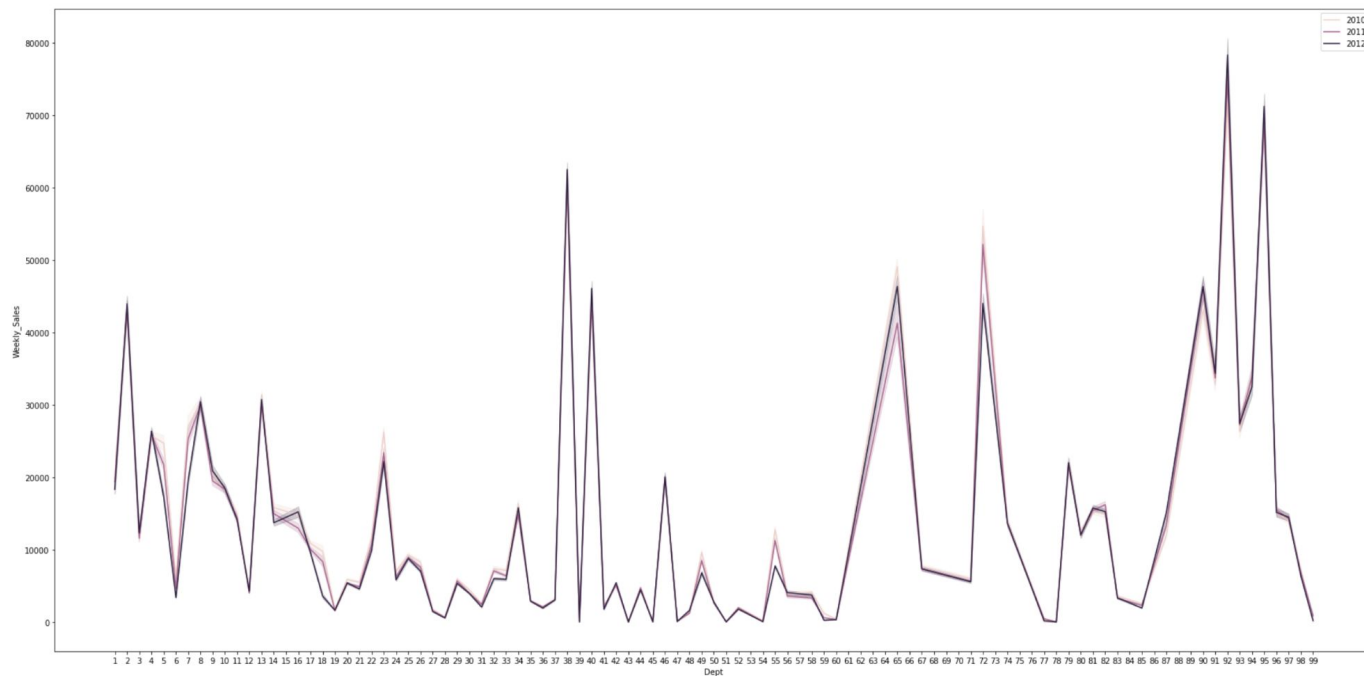
2.EDA(週と売上の関係)

- 11月と12月が1年で売上が最も高くなっている。11月はThanksgiving、12月はクリスマスの前の週となっている。また、2010年より2011年の方が売上が少ない。



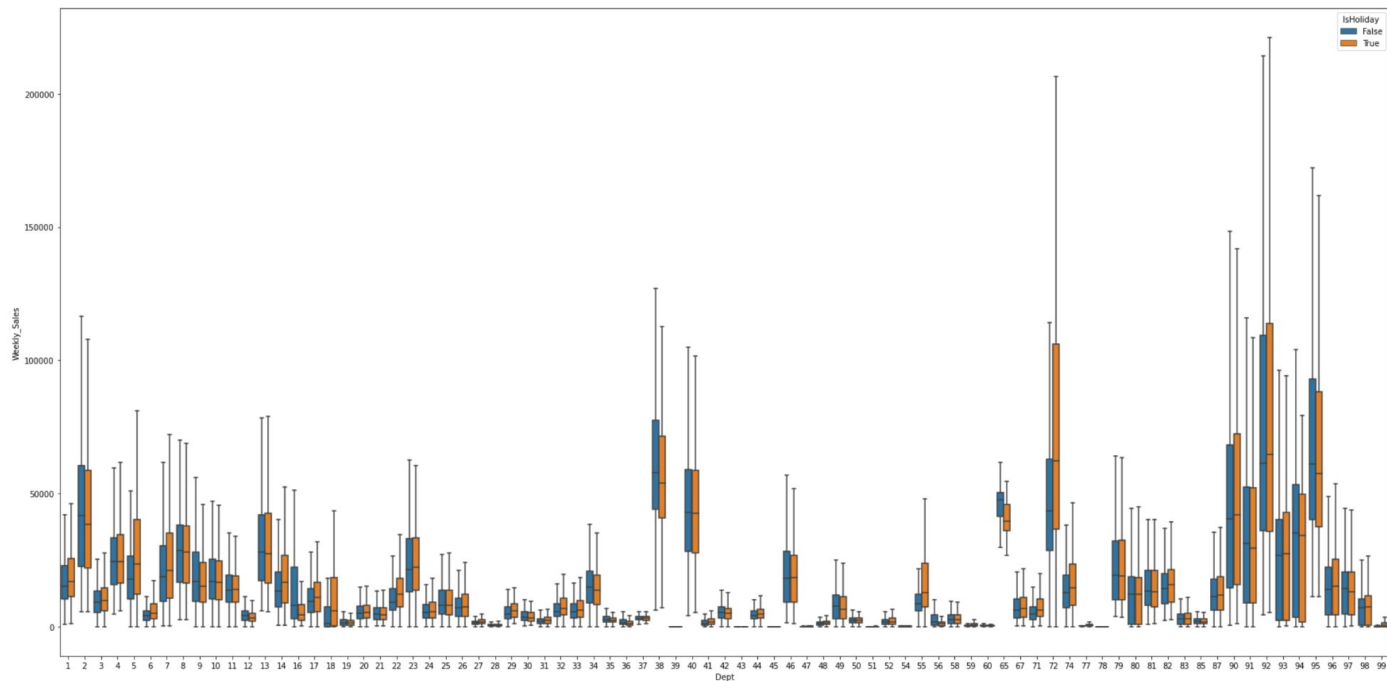
2.EDA(deptと平均売上の関係)

- deptごとに売上の平均は異なり、予測にはdeptが大きく影響することが考えられる。



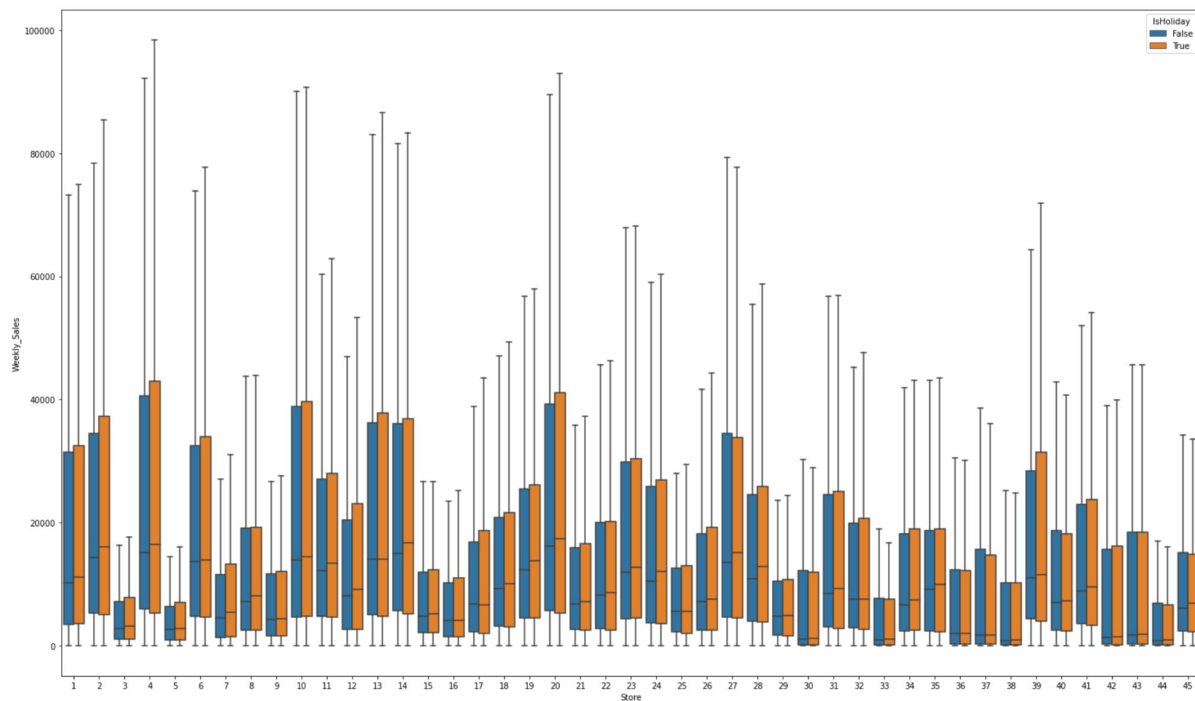
2.EDA(deptと平休日別平均売上の関係)

- dept72は、平日と特別休日で差があるように見えるが、その他のdeptでは平日と特別休日では大きく差がない。



2.EDA(店舗と平休日別平均売上の関係)

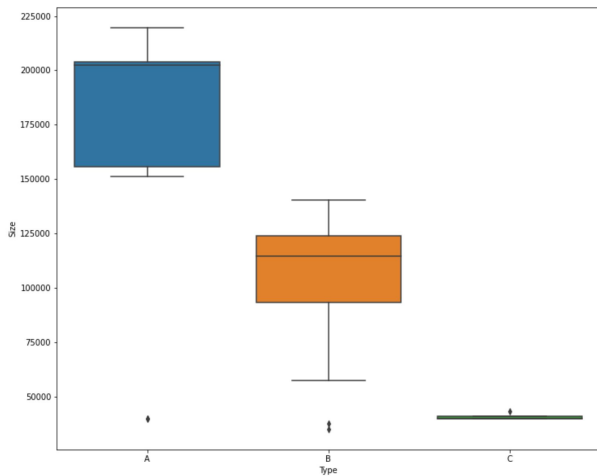
- 店舗別にみると若干ではあるが、どの店舗も特別休日の方が売上が高い傾向となっている。



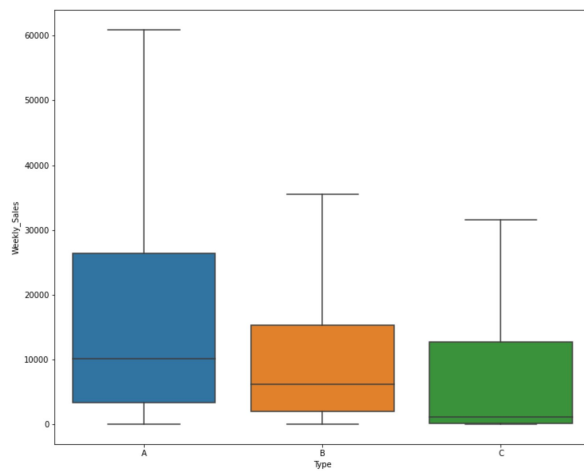
2.EDA(週と売上の関係)

- ・ ショップのタイプ別で店舗の規模が変わっており、売上は店舗規模の大きい店舗が大きくなっている。

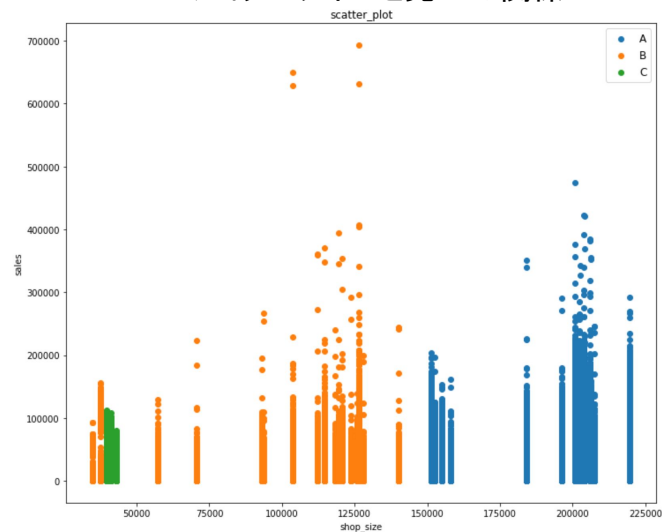
タイプ別ショップサイズの関係



タイプ別週売上の関係



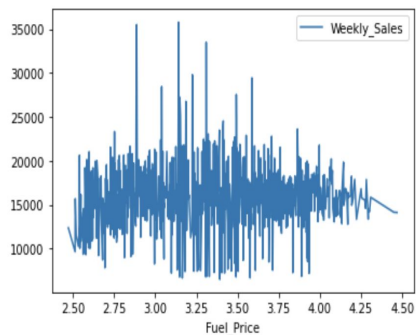
ショップサイズと売上の関係



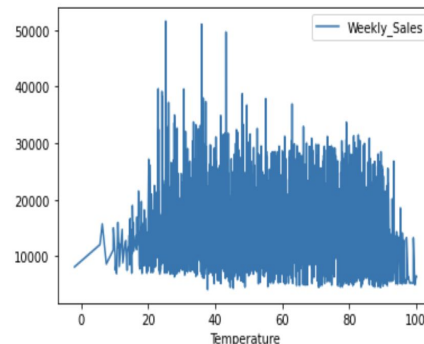
2.EDA(燃料価格、気温、CPI、失業率)

- 燃料価格、気温、CPI、失業率は、週別売上と相関関係が見られず、今回の分析の特徴量から外すこととした。

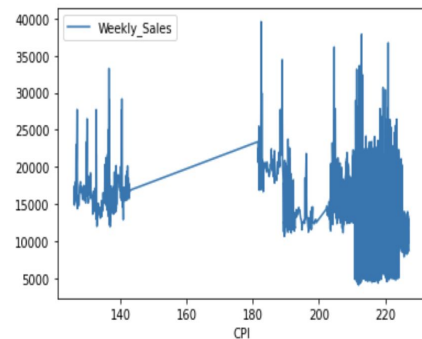
燃料価格



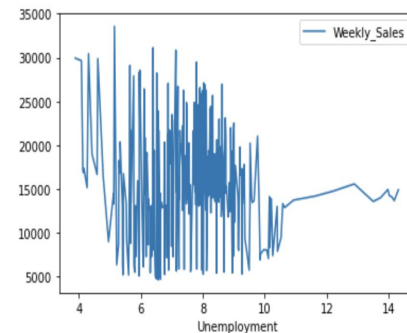
気温



CPI



失業率



3.評価指標

- 今回の評価指標は、WMAEであり、special holidayの重みが通常より5倍の重みとして扱われているため、special holidayの予測の正確さが評価を左右する可能性がある。

This competition is evaluated on the weighted mean absolute error (WMAE):

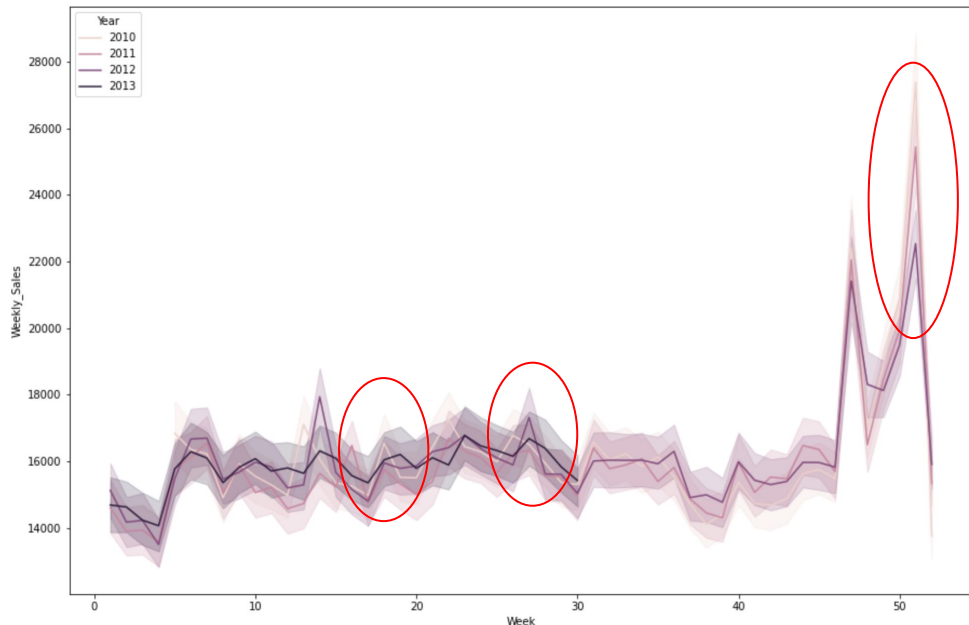
$$\text{WMAE} = \frac{1}{\sum w_i} \sum_{i=1}^n w_i |y_i - \hat{y}_i|$$

where

- n is the number of rows
- \hat{y}_i is the predicted sales
- y_i is the actual sales
- w_i are weights. $w = 5$ if the week is a holiday week, 1 otherwise

4.モデルの考察

- 今回作成のモデルでは、大まかなトレンドを捉えているものの、特別休日のような売上が大きくなるポイントでのズレが大きいことが見て取れるため、特別休日を中心に有効な特徴量の作成や補正が必要



- 使用したモデル
ランダムフォレスト
- 今回の精度
WMAE:3142.07606
- 今後の課題
 - クリスマスの1週間前については、曜日や前後の週との関係調査
 - 売上の大きい特別休日のズレが評価指標にダイレクトに効いてくるため、前年からの伸び率等の特徴量を追加し、検討