

# **PROJEKTOWANIE HURTOWNI DANYCH: EKSPLORACJA DANYCH**

Wiktor Bechciński, 229840

Kamil Budzyn 229850

## 1. CEL BIZNESOWY EKSPLORACJI DANYCH

Zdecydowaliśmy że celem naszej eksploracji biznesowej będzie predykcja sprzedaży rowerów w zależności od grupy wiekowej klientów. Do badania podzieliłismy klientów na cztery grupy: młodzież (<25 lat), młodzi dorośli(25-34 lat), dorośli(35-64 lat) oraz seniorzy(>64lat). Informacje które zdobędziemy pozwolą nam lepiej oszacować na jakich grupach wiekowych należy skupić uwagę w celu zwiększenia zysków.

## 2. TYP PREDYKCJI

Do przedstawienia predykcji wykorzystaliśmy narzędzie Data Mining z programu Visual Studio, predykcję wykonywaliśmy na podstawie wcześniej przygotowanego zbioru danych. Zgłębianie danych (data mining) to proces analityczny, przeznaczony do badania dużych zasobów danych (zazwyczaj powiązanych z zagadnieniami gospodarczymi lub rynkowymi) w poszukiwaniu regularnych wzorców oraz systematycznych współzależności pomiędzy zmiennymi, a następnie do oceny wyników poprzez zastosowanie wykrytych wzorców do nowych podzbiorów danych. Finalnym celem data mining jest najczęściej przewidywanie (zachowań klientów, wielkości sprzedaży, prawdopodobieństwa utraty klienta itp.), dlatego też predykcyjny data mining jest bardzo popularny. Predykcyjny data mining daje bezpośrednie korzyści biznesowe. Proces ten składa się z trzech zasadniczych etapów: (1) wstępnej eksploracji, (2) budowania modelu (z określaniem wzorców) oraz oceny i weryfikacji oraz (3) wdrożenia i stosowania modeli (ang. deployment) dla nowych danych, w celu uzyskania przewidywanych wartości lub klasyfikacji.

## 3. WYBRANA METODA

Do wykonania eksploracji użyliśmy metody Time Series. W odróżnieniu od analiz wykonywanych na próbach losowych, z którymi spotykamy się w przypadku omawiania większości innych statystyk, analiza szeregów czasowych opiera się na założeniu, że kolejne wartości w zbiorze danych reprezentują kolejne pomiary wykonane w równych odstępach czasu. Analiza szeregów czasowych ma dwa główne cele: wykrywanie natury zjawiska reprezentowanego przez sekwencję obserwacji i prognozowanie (przewidywanie przyszłych wartości szeregu czasowego). W tym przypadku metoda Time Series w Visual Studio wykorzystała tryb mieszany korzystając jednocześnie z dwóch algorytmów:

-ARIMA - pozwala na utworzenie modelu autoregresyjnej zintegrowanej średniej ruchomej (ARIMA) umożliwiającego dokładne modelowanie szeregów czasowych. Modele ARIMA oferują bardziej wyrafinowane metody modelowania składników trendu i sezonowości niż modele wygładzania wykładniczego oraz mają dodatkową zaletę polegającą

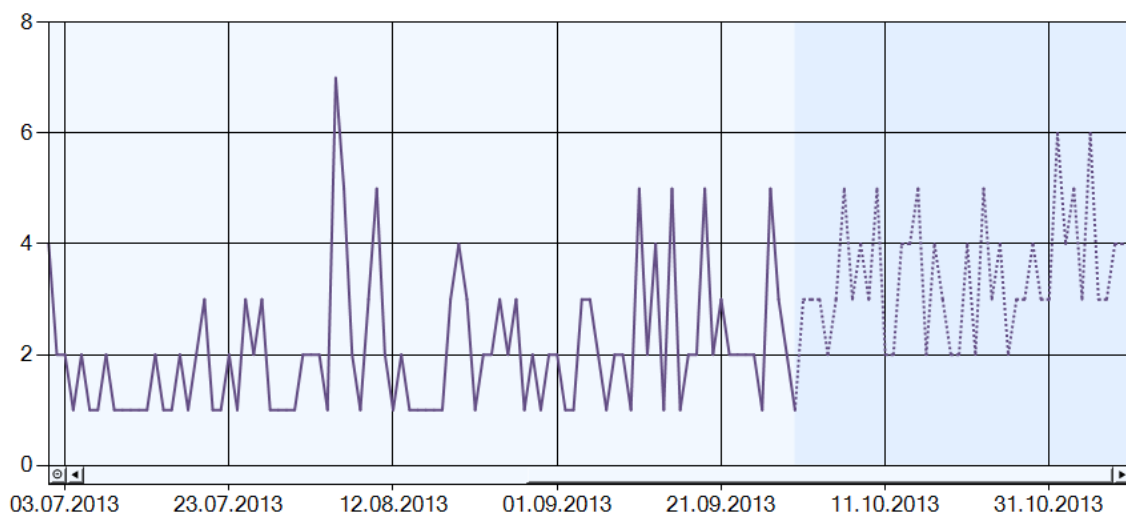
na możliwości uwzględnienia w modelu zmiennych predyktorów  
Metoda pozwala na dostosowanie modelu, określając kolejność autoregresji, różnicowania i średniej kroczącej, jak również sezonowych odpowiedników tych składników.

-ARTXP - autoregresyjny model drzewa do reprezentacji danych okresowych szeregów czasowych. Algorytm ten wiąże zmienną liczbę przeszłych pozycji z każdą bieżącą pozycją, która jest przewidywana. Nazwa ARTXP wywodzi się z faktu, że metoda drzewa autoregresyjnego (algorytm ART) jest stosowana do wielu nieznanych stanów wcześniejszych

## 4. WYKONANE ANALIZY

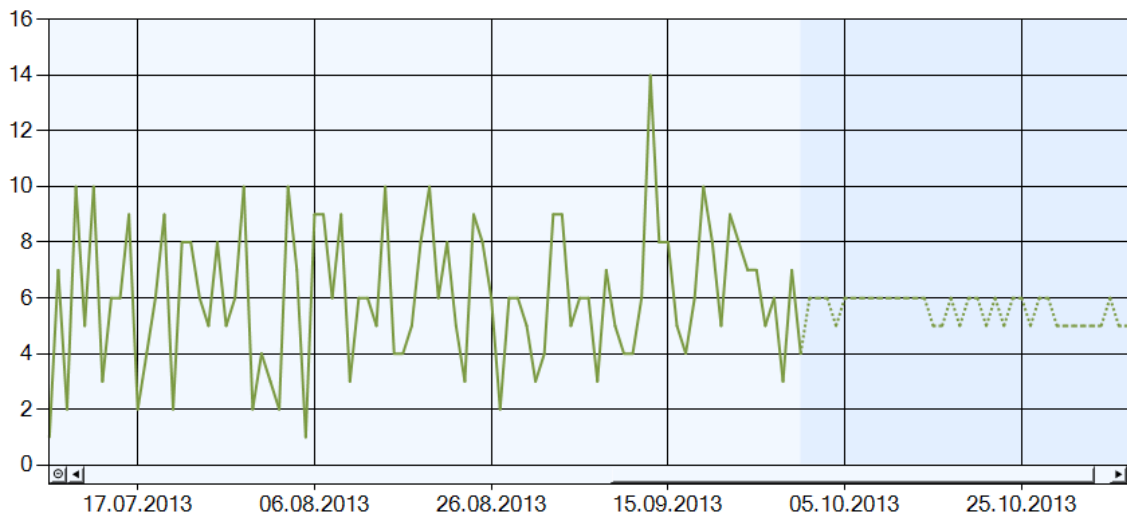
Jakiej ilości zakupionych sztuk możemy się spodziewać w zależności od grupy wiekowej na przestrzeni najbliższych tygodni?

**Wykres 1: Grupa wiekowa młodzież.**



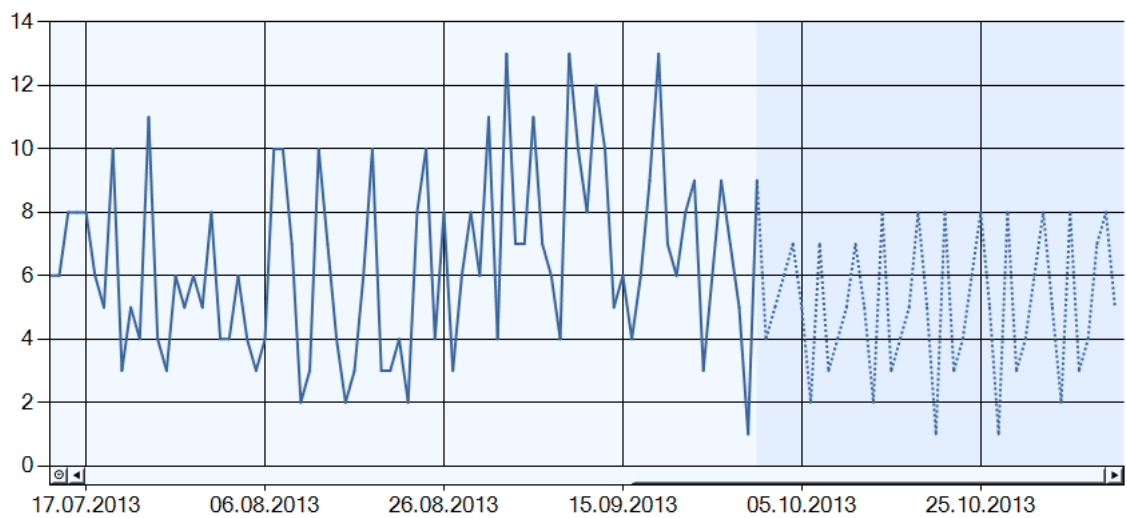
Możemy zauważyć, iż liczba zakupionych rowerów ma przewidywany trend wzrostowy, wartość ta widocznie się podnosi, może to sugerować, iż warto zainteresować się mocniej podaną grupą wiekową w celu zwiększenia sprzedaży.

**Wykres 2: Grupa wiekowa młodzi dorośli.**



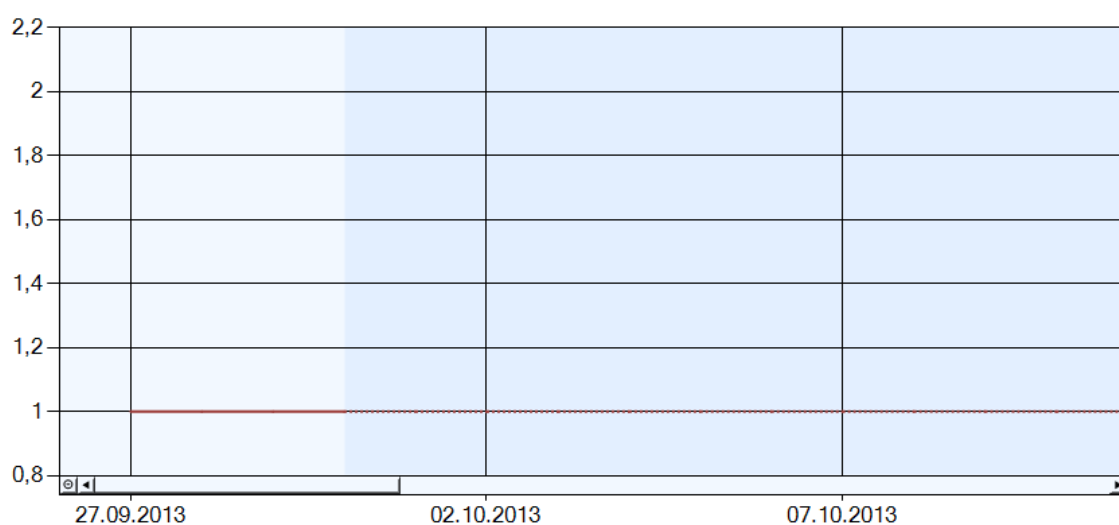
Dla grupy wiekowej młodzi dorośli możemy zauważyć lekki spadek lecz w porównaniu do poprzedniej grupy jest i to i tak wartość wyższa.

**Wykres 3: Grupa wiekowa dorośli.**



Dla tej grupy wiekowej w porównaniu do poprzednich miesięcy możemy zauważyć utrzymanie trendu w średniej ilości 4-6 sztuk na dzień.

**Wykres 4: Grupa wiekowa seniorzy.**



Seniorzy nie składali wystarczająco dużo zamówień aby wykres i predykcja się udała. Lecz nie jest to bezużyteczna wiedza, może to implikować iż sklep powinien poświęcić mniej uwagi tej grupie wiekowej.

## 5. WNIOSKI

Dla grup wiekowych młodzież, młodzi dorośli oraz dorośli narzędzie poradziło sobie dobrze, wykresy poprawnie obrazują możliwe zdarzenia na najbliższe kilka tygodni. W przypadku grupy seniorów narzędzie miało zdecydowanie zbyt mało danych w wyniku czego nie dało się niczego oszacować. Narzędzie nieźle radzi sobie z przewidywaniem krótkiego okresu czasu. Im większa jest liczba danych tym narzędzie radzi sobie lepiej. Technologia może być przydatna w analizie danych.