Dzień dobry. Nazywam się Jakub Walczak i pokrótce zaprezentuję państwu moją znalizę zbioru danych opisanego na platformie kaggle jako „Zbiór danych dotyczący sprzedaży detalicznej globalnego supermarketu z 4 lat.

Jednakże w rzeczywistości mamy do czynienia nie z siecią globalną, a siecią działającą w obrębie USA, ponieważ z tego kraju pochodzą wszyscy klienci. Jednocześnie jak wynika z danych raczej nie ze sprzedażą detaliczną, a ecommence, ponieważ przy każdej transakcji wyszczególniono termin zamówienia i jego wysyłki do klienta.

Pomijając te nieścisłości przejdźmy już do samej analizy.

Na pierwszym z 3 dashbordów pozwoliłem sobie przygotował podsumowanie sprzedaży naszej firmy, aby lepiej zrozumieć jej strukturę i działalność.

W przeciągu 4 lat działalności sprzedaliśmy produkty o łącznej wartości przekraczającej 2 miliony dolarów.

Jeśli chodzi o rozwój naszej firmy perspektywy są bardzo optymistyczne, bowiem z roku na rok sprzedaż wzrasta. Jak można zauważyć na wykresie liniowym, dochodzi jednak do cyklicznych spadków w sprzedaży, które postaram się wyjaśnić na kolejnym dashboardzie.

Przyjżyjmy się jeszcze jednak najpierw strukturze naszych klientów. Z mojej analizy wynika, że większość naszych klientów stanowią konsumenci, jednakże dodając do klientów korporacyjnych grupę home office wówczas zrówna się ona ze zwykłymi konsumentami. Tak więc podczas dalszego prowadzenia naszej działalności nie powinniśmy faworyzować żadnej z tych grup.

Jak już wspominałem nasi klienci znajdują się w całych Stanach. Jak jednak widzimy na poniższym wykresie nasza sprzedaż ma przewagą w zachodniego i wschodnim regionie, co wynika oczywiście ze szczególnego zaludnienia wybrzeży USA. Przygotowałem dla państwa strukturę hierarchiczną, dzięki czemu możemy zajrzeć głębiej i podejrzeć sprzedaż w poszczególnych stanach.

Jak państwo widzicie wśród stanów króluje Kalifornia, co nie dziwi, bowiem w jej obrębie znajdują się takie milionowe miasta jak Los Angeles, San Diego i San Hose, z kolei gdy spojrzymy jeszcze głębiej, na pojedyncze miasto, zdecydowanie największą sprzedaż odnotowano w Nowym Jorku, który jest zdecydowanie najbardziej zaludnionym miastem w USA.

Jeśli chodzi o sprzedaż poszczególnych produktów. Należy zauważyć, że rozkłada się ona w miarę równomiernie między 3 oferowane przez nas kategorie tj. technologię, umeblowanie i artykuły biurowe, z nieznaczną przewagą technologi. Gdy jednak spojrzymy na liczbę transakcji kolejność będzie zupełnie odwrotna. Wynika to oczywiście z faktu, że artykuły biurowe są najtańszą kategorią w naszej ofercie, dlatego sprzedają się najliczniej, generując jednak przy tym najmniejsze wpływy.

Przechodząc dalej należy zauważyć niepokojący cykl, a mianowicie na początku każdego roku nasza sprzedaż spada o ponad 50k dolarów. Spoglądając na wykres kaskadowy wyraźnie widzimy, że sprzedaż ta rośnie w każdym kwartale za wyjątkiem pierwszego kwartału każdego roku, w którym to wyraźnie spada.

Przyjżyjmy się więc tylko zmianom w sprzedaży jakie nastąpiły w pierwszych 3 miesiącach każdego roku. Oczywiście bez roku 2015, który stanowi punkt wyjścia. Na tym wykresie kolumnowym wyraźnie widzimy, że problematycznym miesiącem jest styczeń, który każdorazowo generuje spadek sprzedaży o ponad 50 tysięcy względem grudnia. Jak widać na przygotowanych przeze mnie wskaźnikach, średni spadek to ponad 53 tysiące, a średni spadek w procentach to blisko 70%. Ta liczba obrazuje najlepiej z jak dużym problemem dla naszej firmy mamy do czynienia.

Poszukując już konkretnego miejsca tego spadku należy zauważyć, że w czołówce plasują się najbardziej zaludnione stany, w których jak wspominałem wscześniej mamy największą sprzedaż, a więc California i Nowy Jork. Należy jednak zauważyć, że statystycznie te spadki nie odstają od innych regionów, bowiem gdy spojrzymy na spadki procentowe są one w okolicach średniej i wyników pozostałych regionów.

Warto sprawdzić więc inne podziały, które mogą być przyczyną.

Niniejszy dashboard prezentuje główne segmenty jakie możemy wyróżnić z dostępnego zbioru danych. Podział przez region, przez kategorię produktu i przez segment klienta. Niestety gdy spojrzymy na spadki procentowe między grudniem a styczniem, proste tych spadków w poszczególnych latach dla każdego z podziałów przecinają się. Oznacza to więc, że nie ma takiej kategorii w obrębie danej zmiennej, która z roku na rok ciągnęłaby naszą sprzedaż w styczniu w dół. Spójrzmy dla przykładu na kategorię produktu. W styczniu 2016 roku doszło do najmniejszego procentowego spadku na umeblowaniu, a w styczniu 2018 zdecydowanie najmniejszy spadek był na technologii. Ten sam brak korelacji spadku z kategorią widzimy w pozostałych zmiennych.

Prowadzi to do wniosku, że w naszym zbiorze danych jedyną zmienną objaśniającą spadek pozostaje czas – a dokłądnie miesiąc.

Szukając przyczyny w samym miesiącu, pierwszym wyjaśnieniem jakie przychodzi na myśl są święta bożego narodzenia, które mogą zwiększać sprzedaż w grudniu i tym samym powodować spadek w styczniu. Jednakże z danych wynika, że nie dochodzi do istotnego wzrostu sprzedaży w naszej firmie w grudniu, więc święta nie mogą być głównym powodem spadku w styczniu. Nie zamieściłem takiego wykresu na dashboardzie z braku miejsca, ale grudniowa sprzedaż była przeze mnie sprawdzana.

Powodu spadku w styczniu należy szukać więc gdzieś indziej, innym prawdopodobnym powodem mogą być problemy z dostawami do naszych sklepów, co powoduje, że nasze zapasy, a więc i podaż są mniejsze w styczniu i nie wystarczające względem popytu klienta przez co sprzedaż maleje. Powinniśmy, więc przeanalizować nasze łańcuchy dostaw i ustalić, czy różnią się one w styczniu w sposób istotny względem reszty roku (niestety, ten zestaw danych nie oferuje takich informacji).

Na koniec chciałbym jeszcze dodać, że dodatkowym powodem dla spadku w segmencie korporacyjnym może być również planowanie budżetów na nowy rok przez naszych klientów i tym samym dokonywanie większości zakupów dopiero od lutego. To wytłumaczenie nie tłumaczy jednak segmentu zwykłych konsumentów.

Jednocześnie chciałbym przeprosić za przekroczenie czasu i prosić o zignorowanie opisu, który zamieściłem w prawym dolnym rogu ostatniej strony. Zrobiłem to dlatego, że projekt ten będę wrzucał później na githuba i bez tego opisu zabrakłoby wniosków z analizy, ponieważ nie znajdzie się tam nagranie głosowe.

To wszystko. Dziękuję za uwagę.