

PROJEKT

Analiza i Klasyfikacja Danych (Wersja Uproszczona)

Grupa: Poniedziałek TN 9:15-11:00

Autor:
Damian Kędzierski 260493

Prowadzący: dr inż. Marek Lubicz

Spis treści

.....	3
Założenia Projektu	3
<i>Pulpit Menadżerski – Wskazówki Metodyczne.....</i>	<i>3</i>
OMÓWIENIE ZADANIA – Wersja Uproszczona.....	4
Zrozumienie uwarunkowań biznesowych (P2).....	5
1. Definiowanie przekazu:.....	5
2. Określanie odbiorców:	5
3. Określanie wskaźników wydajności (KPI):.....	5
4. Lista potrzebnych źródeł danych:.....	6
5. Określanie wymiarów i filtrów:	7
6. Drążenie danych:.....	7
7. Harmonogram aktualizacji:	7
<i>Projekt układu i rozmieszczenia elementów.....</i>	<i>7</i>
Zrozumienie danych (P2)	8
Przygotowanie Danych (P2)	9
1. Czyszczenie danych (niepoprawnych, brakujących, nieistotnych informacji).....	9
2. Integracja danych (Power Query & Power Pivot)	9
3. Selekcja danych	10
4. Transformacja danych	10
5. Eksploracja danych – zbiór ogólnych opisów.....	11
Sprawozdanie część techniczna (TABLEAU) – P3.....	13
<i>Modelowanie Pulpit menadżerskiego</i>	<i>13</i>
Sprawozdanie część dla decydenta – P4.....	16
<i>Ewaluacja i wdrożenie</i>	<i>16</i>
<i>STORY – Analiza przedsiębiorstwa</i>	<i>16</i>
<i>Podsumowanie</i>	<i>19</i>

ZAŁOŻENIA PROJEKTU

- Przedmiotem projektu jest rozwiązanie standardowego **problemu biznesowego** dla wygenerowanych danych biznesowych, różnych od danych wykorzystywanych na zajęciach, z wykorzystaniem wybranych technik **analityki deskryptywnej** i prostej **analityki predykcyjnej** w **Tableau Desktop**.
- Oczekiwanym **najważniejszym wynikiem projektu** nie jest budowa jednego lub kilku pulpitów menedżerskich (zestawów wizualizacji na danych biznesowych), ale uzyskanie wyników, które można uznać za rozwiązanie problemu biznesowego lub wskazówki dla decydenta odnośnie takiego rozwiązania
- **dane biznesowe** powinny odnosić się do ustalonej **sytuacji biznesowej** w dowolnej dziedzinie, w której występują określone
 - obiekty (np. klienci, pracownicy, firmy, programy),
 - oznaczone charakterystyką kategorii (np. terytorialną, czasową, zawodową, rodzajową)
 - zdarzenia dotyczące tych obiektów oznaczone wymiarem czasu (np. sprzedaż, migracja, inwestycje, kondycja finansowa)
- dane biznesowe powinny być **unikalne**
- dane biznesowe można znaleźć np. na stronach [Kaggle](#), [github](#), [KDNuggets](#) i w innych źródłach podanych w zakładce [Źródła data science](#)
- Opracowane pulpity jednoznacznie związane z wybranymi celami biznesowymi

Pulpit Menadżerski – Wskazówki Metodyczne

- Opracowane pulpity powinny być jednoznacznie związane z wybranymi **celami biznesowymi** i zgodne z projektem.
- Pulpity powinny być **czytelne** (nie więcej niż 4-8 wizualizacji, zależnie od wielkości i szczegółowości).
 - Proszę ustawić wielkość pulpitu w **Tableau (Size)** na Automatic oraz rodzaj pulpitu na Default (analogicznie w innych środowiskach analitycznych).
- Wizualizacje powinny pokazywać:
 - **dynamikę wybranych zjawisk** (zmienność w czasie, trendy, tempo zmian) z uwzględnieniem wniosków (opcjonalnie prognozy) dla kolejnych okresów
 - oraz w odniesieniu do ustalonych, wybranych przez autorów **grup obiektów** (kategorii obiektów):
 - produktów (hierarchia!),
 - obszarów (hierarchia!),
 - okresów kalendarzowych, itp.
- Wszystkie wielkości wykorzystywane do tworzenia wizualizacji powinny być wyliczane na podstawie danych źródłowych, a nie wklejane przez wartości.
 - Dopuszczalne wartości wskaźników **KPI** (np. wykorzystywane w definiowaniu przedziałów oceny (wynik1 od do, wynik2 od do,...) można zdefiniować jako
 - parametry (wyższa ocena punktowa)
 - lub wkleić jako konkretne liczby (niższa ocena punktowa).
- W pulpitych wykonanych w **Tableau** można zastosować poznane na zajęciach funkcjonalności:
 - mapy lub wykresy dual axis, top5/bottom5, target% (ustalony przez decydenta, a nie odczytany z danych),
 - możliwość pokazywania bardziej lub mniej szczegółowych danych (np. dotyczących różnych poziomów hierarchii czasowej, produktowej, geograficznej).
 - Możesz również zastosować następujące funkcjonalności (będą omawiane w zadaniach 4-6):
- **dynamiczne wizualizacje** z polami wyboru perspektywy (np. miary KPI),
- analizy **what-if** (plan analizy, opis wykonania, ocena wyników zmian ustalonych wielkości na miary KPI),
- **modele grupowania i prognozowania** (zakładka Analytics)
- **trendy** i proste **modele predykcyjne**
- Uporządkuj plik wynikowy, sprawdź nazwy zakładek i zmiennych (powinny być jednoznaczne),
 - sprawdź filtry/fragmentatory (unikalne do wszystkich wizualizacji na pulpicie), **czytelne** tytuły wszystkich wizualizacji.

OMÓWIENIE ZADANIA – WERSJA UPROSZCZONA

- ⇒ Masz zlecenie od KRAJOWEGO centrum sprzedaży FIRMY na wykonanie aplikacji (pulpit lub zestaw powiązanych pulpitów menedżerskich), wspomagającej **analizę danych sprzedażowych** dla potrzeb ZARZĄDU lub REGIONALNYCH CENTRÓW SPRZEDAŻY.
- ⇒ Zarząd wymaga udostępnienia informacji strategicznych zagregowanych, dotyczących całego kraju, przydatnych do podejmowania **decyzji dotyczących przyszłych działań** i opcjonalnej możliwości **bardziej szczegółowej analizy konkretnego regionu** tylko w wyjątkowych przypadkach (np. załamanie sprzedaży określonej kategorii produktów w ostatnim kwartale) oraz **wglądu do sytuacji finansowej** Firmy (nie pojedynczych zamówień).
- ⇒ **Konsultanci Regionalnych Centrów Sprzedaży** powinni mieć możliwości bieżących charakterystyk sprzedaży w poszczególnych miastach i województwach w jednym wybranym regionie (aplikacja powinna być uniwersalna, możliwa do zastosowania w każdym regionie).
- Opracuj **konceptję** aplikacji (pulpitu menedżerskiego i/lub zestawu pulpitów), realizującej wskazane cele.
 - Wskaż wybraną przez Ciebie **grupę użytkowników** aplikacji oraz **wybrane środowisko analityczne**.
 - Opracuj **projekt** pulpitu (pulpitów), zgodnie z metodyką przedstawioną na wykładzie.
 - Zdefiniowanie celów biznesowych, analitycznych i KPI jest obowiązkowe (warunek konieczny otrzymania punktów).
 - Określenie np. kwota zakupu lub czas rozmowy jako KPI nie jest wystarczające.
 - Projekt zapisz w pliku projekt_numerindeksu.pdf.

Koncepcja Procesu Analitycznego



ZROZUMIENIE UWARUNKOWAŃ BIZNESOWYCH (P2)

1. Definiowanie przekazu:

Pulpit menadżerski powinien w prosty i efektywny sposób ukazać graficzny interfejs prezentujący dane w ujęciu całościowym za pomocą odpowiednich wskaźników i z uwzględnieniem poszczególnych aspektów lub procesów biznesowych.

⇒ Tworząc pulpit dla firmy, głównym celem będzie **maksymalizacja zysku, przy jak najmniejszych nakładach (kosztach)**

Problemem biznesowym jest **ocena sprzedaży produktów**. Celem oceny sprzedaży produktów jest identyfikacja najbardziej rentownych kategorii produktów, a więc należy ustalić:

- Jaką kategorię produktów należy traktować jako kluczową i skupić na niej działania przedstawicieli? Jaka kategoria produktów jest naszym głównym źródłem przychodu?
- Jakie kategorie produktów posiadają zamówienia na najwyższe kwoty?
- Jakie kategorie produktów cieszą się największym powodzeniem?
- Jaki jest najczęściej priorytet zamówienia wśród poszczególnych segmentów rynku?
- Czy wśród zamówień konkretnych kategorii produktów można zaobserwować sezonowość?

2. Określanie odbiorców:

- Końcowymi użytkownikami pulpitu będą konsultanci regionalnych centrów sprzedaży i zarząd
 - powinni mieć możliwości bieżących charakterystyk sprzedaży w poszczególnych miastach i województwach w jednym wybranym regionie (aplikacja powinna być uniwersalna, możliwa do zastosowania w każdym regionie)
- Dzięki temu pulpitowi firma, będzie efektywniej planować przyszłe działania oraz w wyjątkowych sytuacjach będzie bardziej szczegółowo analizować konkretne regiony.

3. Określanie wskaźników wydajności (KPI):

KPI = miary, na które nakładane są określone przez firmę cele.

- Wszystkie poniższe wskaźniki będą odnosić się do rodzaju (kategorii) produktów sprzedawanych konkretnym klientom (segment rynku) dotyczących całego kraju (regiony, województwa, miasta i powiaty)

Wskaźnik efektywności sprzedaży podkategorii produktów w regionach

Wskaźnik efektywności sprzedaży to miarodajna metryka, która pozwala ocenić skuteczność działań sprzedażowych przedsiębiorstwa.

Interpretacja

- Wartość wskaźnika < 70% średniej sprzedaży → wartość niezadowalająca
- Wartość wskaźnika pomiędzy 70%, a 110% średniej sprzedaży → wartość standardowa
- Wartość wskaźnika > 110% średniej sprzedaży → wartość dobra

Roczny Wzrost Sprzedaży

Znając tę miarę możemy jednoznacznie określić, czy uzyskany wynik pozwala firmie się rozwijać. Miara ta jest szczególnie istotna, ponieważ wzrost sprzedaży jest kluczowy dla przetrwania firmy.

$$RWS = \frac{\text{Przychód w roku badanym} - \text{Przychód w roku referencyjnym}}{\text{Przychód w roku badanym}}$$

Interpretacja

- Jeśli sprzedaż w roku badanym jest większa od sprzedaży w roku referencyjnym to jest to wartość dobra - pożądana (kolor zielony).
 - $RWS > 0 \rightarrow$ wartość pożądana
- Jeśli sprzedaż w roku badanym jest mniejsza od sprzedaży w roku referencyjnym to jest to wartość zła - niepożądana (kolor czerwony).
 - $RWS < 0 \rightarrow$ wartość niepożądana

Średnia rentowność sprzedaży

Wskaźnik informuje o tym ile procent sprzedaży stanowi marża zysku po odliczeniu wszystkich kosztów. Wyższy poziom tego wskaźnika wskazuje na korzystniejszą kondycję finansową przedsiębiorstwa.

$$\acute{S}RS = \frac{\text{Zysk}}{\text{Przychody}} \times 100\%$$

Interpretacja \rightarrow główną miarą interpretacyjną wpływającą na wskaźnik jest zysk, który jest indywidualnie definiowany przez odbiorców tkz. próg zadowolenia osiągniętego zysku

- Jeśli zysk $>$ próg zadowolenia \rightarrow wartość pożądana (zielony)
- Jeśli zysk $<$ próg zadowolenia \rightarrow wartość niepożądana (czerwony)

Średnia Wielkość Zakupu

Wskaźnik pozwala dowiedzieć się jaką kwotę statystycznie zostawia u nas klient. Sukcesywne rozwijanie procesów sprzedaży w firmie może w znaczny sposób wpłynąć na wynik firmy.

$$\acute{S}WZ = \frac{\text{Przychód}}{\text{Ilość sprzedaży}}$$

4. Lista potrzebnych źródeł danych:

- Do przygotowania pulpitu niezbędnie potrzebne będą historyczne dane sprzedażowe wraz z danymi o regionach oraz ewentualnie o powiatach, które powinny zostać zagregowane. Surowe dane znajdują się w pliku „Data-Generator AKD_260493.xlsm”.
- Dane shp o powiatach oraz dane z kodami powiatów
- Dane powinny być aktualizowane na bieżąco, aby mieć możliwość podejmowania decyzji dotyczących przyszłych działań.

5. Określanie wymiarów i filtrów:

Wymiary

- Okresy → lata, kwartały, miesiące, dni
- Regiony → centralny, południowo-zachodni, południowy, północno-zachodni, północny, wschodni
- Segment rynku → biuro w domu, korporacja, mała firma
- Kategorie produktów → artykuły biurowe, meble, urządzenia techniczne
- Priorytet zamówienia → krytyczny, niski, średni, wysoki

Filtry

- Możliwość wyboru konkretnych kategorii produktów
- Możliwość wyboru okresu
- Możliwość wyboru regionu
- Możliwość wyboru segmentu rynku
- Możliwość wyboru priorytetu zamówienia
- Możliwość wyboru progu zadowolenia z zysku

6. Drażenie danych:

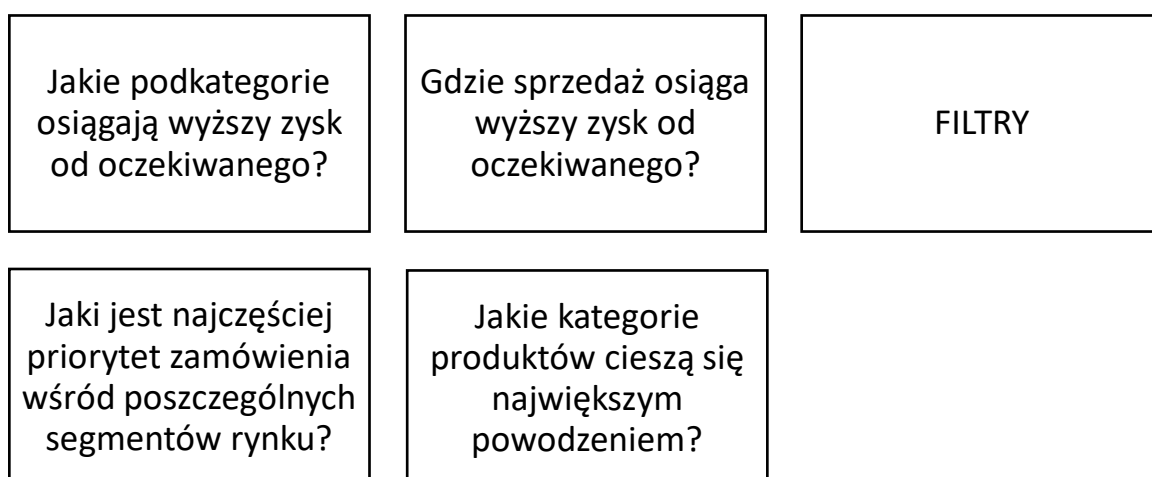
Drażenie danych będzie dotyczyć bardziej szczegółowej analizy konkretnego regionu oraz wglądu do analizy dynamiki sprzedaży produktów i analizy porównawczej sprzedaży.

7. Harmonogram aktualizacji:

Pulpit powinien zostać rozszerzony o nowy wskaźnik KPI skuteczności (obliczany poprzez dzielenie ilości wykorzystanych szans sprzedaży przez ilość wszystkich szans sprzedaży) - do którego brakuje danych o ilości niewykorzystanych szansach sprzedaży. To bardzo dobra miara, która dobrze zarządzana może być kluczem do wzrostu firmy.

- Dane powinny być aktualizowane przynajmniej raz na kwartał.

Projekt układu i rozmieszczenia elementów





ZROZUMIENIE DANYCH (P2)

Historyczne dane sprzedażowe oraz tabela podziału na regiony i powiaty. Wygenerowane i skopiowane z pliku „Data-Generator AKD_260493.xlsm” do pliku „dane_260493.xlsx”.

Skopiowane dane składają się z 3 arkuszy:

- „Baza_danych” – Główny arkusz z danymi. Zawiera dane sprzedażowe.

Nazwa ▾	Typ ▾	Klucz
Województwo	A ^B _C Tekst	
Miasto	A ^B _C Tekst	
Produkt	A ^B _C Tekst	
Segment Rynku	A ^B _C Tekst	
Priorytet zamówienia	A ^B _C Tekst	
Data zamówienia	 Data	
Order ID	1 ² ₃ Pełna liczba	
Data wysyłki	 Data	
Liczba	1 ² ₃ Pełna liczba	
Cena jednostkowa	1 ² ₃ Pełna liczba	
Koszt jednostkowy	1 ² ₃ Pełna liczba	
Wartość	1 ² ₃ Pełna liczba	
Koszty	1 ² ₃ Pełna liczba	
Zysk	1 ² ₃ Pełna liczba	
Kategoria	A ^B _C Tekst	
Podkategoria	A ^B _C Tekst	

- „Regiony” – zawiera informacje o regionach w konkretnych województwach przedstawicieli

Nazwa ▾	Typ ▾	Klucz
Region	A ^B _C Tekst	
Województwo	A ^B _C Tekst	
Przedstawiciel	A ^B _C Tekst	

- „Powiaty” – zawiera informacje jakie miasta i powiaty znajdują się w konkretnym województwie

Nazwa ▾	Typ ▾	Klucz
WOJEWÓDZTWA	A ^B _C Tekst	
MIASTA	A ^B _C Tekst	
POWIATY	A ^B _C Tekst	

Order ID – 4000 unikatowe wartości (zamówienia)

Zmienne kategoryczne (jakościowe)

Nominalne

- Województwo – 16 unikatowych wartości
- Miasto – 367 unikatowych wartości
- Produkt – 827 unikatowych wartości
- Przedstawiciel – 16 unikatowe wartości
- Powiaty – 367 unikatowe wartości
- Region – 6 unikatowe wartości

Uporządkowane

- Segment rynku – 3 unikatowe wartości
- Priorytet zamówienia – 4 unikatowe wartości
- Kategoria – 3 unikatowe wartości
- Podkategorie – 12 unikatowe wartości

Zmienne ilościowe (będą przeanalizowane w dalszej części – eksploracji danych)

Przedziałowe

- Data zamówienia (od 2015 do 2022)
- Data wysyłki (od 2015 do 2022)

Zliczenia i Ilorazowe

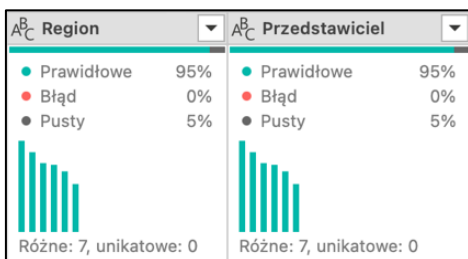
- Liczba
- Cena jednostkowa
- Koszt jednostkowy
- Wartość Koszty
- Zysk

PRZYGOTOWANIE DANYCH (P2)

Eksploracyjna Analiza Danych – przygotowanie danych

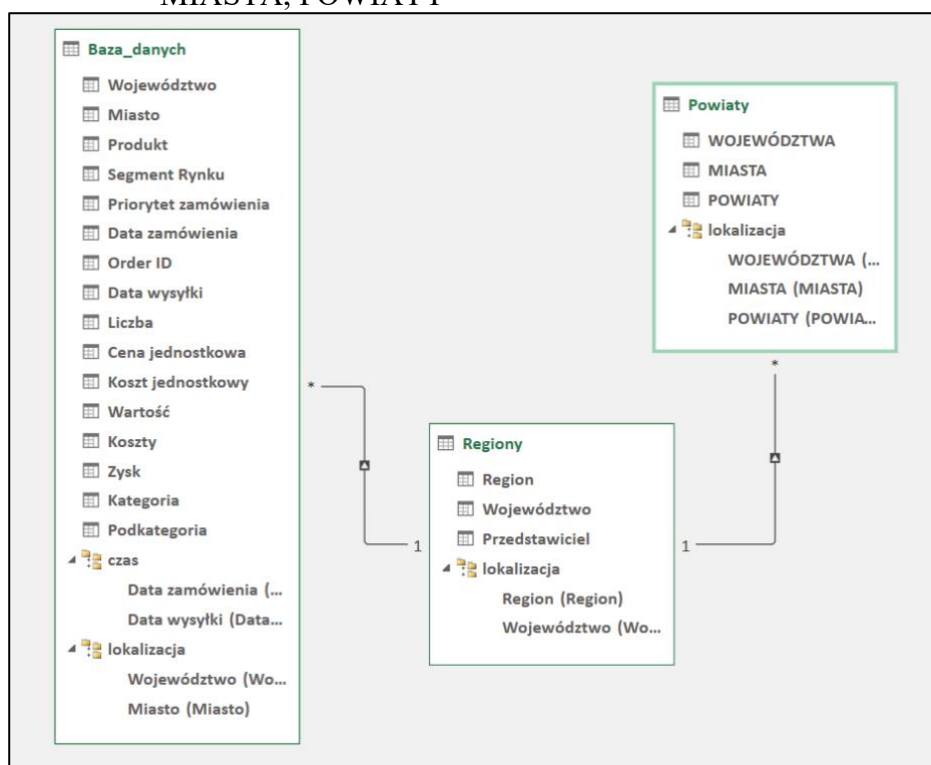
1. Czyszczenie danych (niepoprawnych, brakujących, nieistotnych informacji)

- a. Początkowo dane o województwach scalają się w 95% (202 wierszy nie mogą się scalać), ponieważ z powodu innej nazwy jednej wartości województw w „Baza_danych” – zachodnio-pomorskie i „Regiony” – zachodniopomorskie, występuje błąd.
 - i. ROZWIĄZANIE → Zmiana nazwy województwa zachodniopomorskie w „Regiony” na **zachodnio-pomorskie**



2. Integracja danych (Power Query & Power Pivot)

- a. Power Query → Scalenie *Baza_danych* i *Regiony*, oraz dodatkowo scalenie *Powiaty* z *Regiony* (*Powiaty* nie mogą być scalone bezpośrednio z *Baza_danych*, ponieważ zaszłoby zjawisko tkz. redundancji danych)
 - i. Dane były scalane po województwach.
- b. Power Pivot → Utworzenie relacji między *Baza_danych*, *Regiony* i *Powiaty* (atrybutem relacji były dane o województwach)
 - i. Utworzenie hierarchii czasowej – Data zamówienia, Data Wysyłki
 - ii. Utworzenie hierarchii terytorialnej – REGION, WOJEWÓDZTWA, MIASTA, POWIATY



3. Selekcja danych

tylko istotne dane pozyskane z bazy danych do zadań analitycznych.

4. Transformacja danych

Sprowadzenie do postaci tabel

- Transformacja kolumny cena jednostkowa i koszt jednostkowy w marża jednostkowa (cena jednostkowa – koszt jednostkowy)
- Dodanie kolumny „Rentowność” (Zysk / Wartość)
- Zmiana nazwy kolumny „Wartość” na „Przychód”
- Zmiana typu danych cena jednostkowa, Koszt jednostkowy, Przychód, Koszty, Zysk na typ Waluta

Nazwa 	Typ 	Klucz 
Order ID	1 ² 3 Pełna liczba	
Przedstawiciel	A ^B C Tekst	
Segment Rynku	A ^B C Tekst	
Produkt	A ^B C Tekst	
Kategoria	A ^B C Tekst	
Podkategoria	A ^B C Tekst	
Data zamówienia	 Data	
Priorytet zamówienia	A ^B C Tekst	
Data wysyłki	 Data	
Liczba	1 ² 3 Pełna liczba	
Cena jednostkowa	\$ Waluta	
Koszt jednostkowy	\$ Waluta	
Marża jednostkowa	\$ Waluta	
Przychód	\$ Waluta	
Koszty	\$ Waluta	
Zysk	\$ Waluta	
Rentowność	% Wartość procentowa	
Region	A ^B C Tekst	
Województwo	A ^B C Tekst	
Miasto	A ^B C Tekst	
Powiat	A ^B C Tekst	

5. Eksploracja danych – zbiór ogólnych opisów

Charakterystyki zmiennych ilościowych w środowisku **R** za pomocą funkcji `summary()` oraz funkcji `describe()` z biblioteki `psych` (do głębszej analizy zmiennych ilościowych zliczeniowych i ilorazowych)

`summary()`

Data zamówienia		Data wysyłki	
Min.	:2015-01-01 00:00:00	Min.	:2015-01-04 00:00:00
1st Qu.	:2016-11-23 06:00:00	1st Qu.	:2016-12-19 18:00:00
Median	:2018-11-25 00:00:00	Median	:2018-12-19 00:00:00
Mean	:2018-11-24 00:54:58	Mean	:2018-12-19 01:01:25
3rd Qu.	:2020-11-08 00:00:00	3rd Qu.	:2020-12-05 18:00:00
Max.	:2022-11-11 00:00:00	Max.	:2022-12-28 00:00:00

Liczba	Marża jednostkowa	Wartość	Koszty	Zysk	Rentowność
Min. : 1.0	Min. : 1.00	Min. : 21	Min. : 17	Min. : 4	Min. : 0.1495
1st Qu.: 260.0	1st Qu.: 5.00	1st Qu.: 10990	1st Qu.: 9168	1st Qu.: 1860	1st Qu.: 0.1507
Median : 517.5	Median : 16.00	Median : 39526	Median : 33250	Median : 6219	Median : 0.1532
Mean : 508.9	Mean : 52.93	Mean : 175973	Mean : 149345	Mean : 26628	Mean : 0.1622
3rd Qu.: 757.8	3rd Qu.: 57.00	3rd Qu.: 164852	3rd Qu.: 139862	3rd Qu.: 24908	3rd Qu.: 0.1667
Max. : 1000.0	Max. : 1339.00	Max. : 7943250	Max. : 6751540	Max. : 1191710	Max. : 0.2857

- Min – minimalna wartość
- 1st Qu. – pierwszy kwartył
- Median – Mediana
- Mean – Średnia arytmetyczna
- 3st Qu. – trzeci kwartył
- Max – maksymalna wartość

`describe()`

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
Liczba	1	3798	508.94	288.69	517.50	510.80	368.43	1.00	1000.00	999.00	-0.06	-1.21	4.68
Marża jednostkowa	2	3798	52.93	113.41	16.00	29.67	19.27	1.00	1339.00	1338.00	6.16	50.42	1.84
Wartość	3	3798	175972.71	427132.23	39526.50	88327.58	52343.93	21.00	7943250.00	7943229.00	7.50	83.55	6930.82
Koszty	4	3798	149345.16	363037.55	33249.50	74835.47	44205.94	17.00	6751540.00	6751523.00	7.50	83.56	5890.80
Zysk	5	3798	26627.55	64095.10	6219.00	13493.45	8060.15	4.00	1191710.00	1191706.00	7.49	83.47	1040.03
Rentowność	6	3798	0.16	0.02	0.15	0.16	0.00	0.15	0.29	0.14	3.43	14.37	0.00

- n = ilość w zbiorze danych
- mean – średnia arytmetyczna
- sd (standard deviation) – odchylenie standardowe
- median – mediana
- trimmed – średnia obcięta (z domyślną wartością przycięcia 0,1)
- mad (median absolute deviation) - mediana bezwzględnego odchylenia (od mediany).
- Min – minimalna wartość
- Max – maksymalna wartość
- Range – zasięg wartości (różnica między max i min)
- Skew - Oblicza skośność zbioru danych, który opisuje symetrię tego zbioru danych wokół średniej.
- Kurtosis - Jedna z miar kształtu rozkładu wartości cechy. Kurtosis określa intensywność występowania wartości skrajnych, mierzy więc ona, co się dzieje w "ogonach" rozkładu.
 - Wysoka dodatnia wartość dla kurtozy wskazuje na to, iż krańce rozkładu są dłuższe niż te dla rozkładu normalnego
 - Ujemna wartość dla kurtozy wskazuje na krótsze krańce (podobnie jak w rozkładach prostokątnych).
- Se (standard error) – standardowy błąd średniej

Po wstępnej eksploracji danych na pierwszy rzut oka zauważalne jest, że okres zamieszczonych danych jest od pierwszego kwartału 2015 do czwartego kwartału 2022. Ponadto jest bardzo duże rozproszenie danych, ponieważ najmniejsze zamówienie było tylko na jedną rzecz, natomiast największe na 1000 sztuk, zarówno przychód, którego minimalna wartość wynosiła zaledwie 21 zł, a maksymalna ponad 7 mln zł (Analogicznie z pozostałymi danymi finansowymi – Koszty, Zysk), jednakże w danych finansowych wartości maksymalne są odstające, ponieważ różnica między 3 kwartylem a wartosiami maksymalnymi są ogromne, co świadczy o tym, że bardzo mało było zamówień na tak wysokie kwoty (różnica między kwartylami jest na poziomie kilkudziesięciu tysięcy, a między wartością maksymalną i 3 kwartylem to już różnica ponad 7 mln). Dowodzi również tego skośność zbioru danych wokół średniej (skew), gdzie tylko w Liczba jest na bardzo niskim poziomie, a w pozostałych danych jest zdecydowanie większy, co przekłada się wcześniej wspomniane różnice wartości odstających od średniej. Mediana jest większa od średniej wartości liczby zamówień, przychodów i kosztów, co świadczy o statystycznie mniejszych wartościach od średniej, spowodowane dużą różnicą między wartością maksymalną a pozostałymi. Analizując wskaźnik kurtozy, ujemna wartość występuje tylko w liczbie zamówień co wskazuje na krótsze krańce, a w pozostałych danych są dodatnie, czyli krańce rozkładu są dłuższe niż te dla rozkładu normalnego.

SPRAWOZDANIE CZĘŚĆ TECHNICZNA (TABLEAU) – P3

Na tym etapie tworzy się prototyp pulpitu menadżerskiego, dobiera się odpowiednie wskaźniki i wizualizacje, które będą odpowiednie dla potrzeb menadżerów.

Modelowanie Pulpitu menadżerskiego

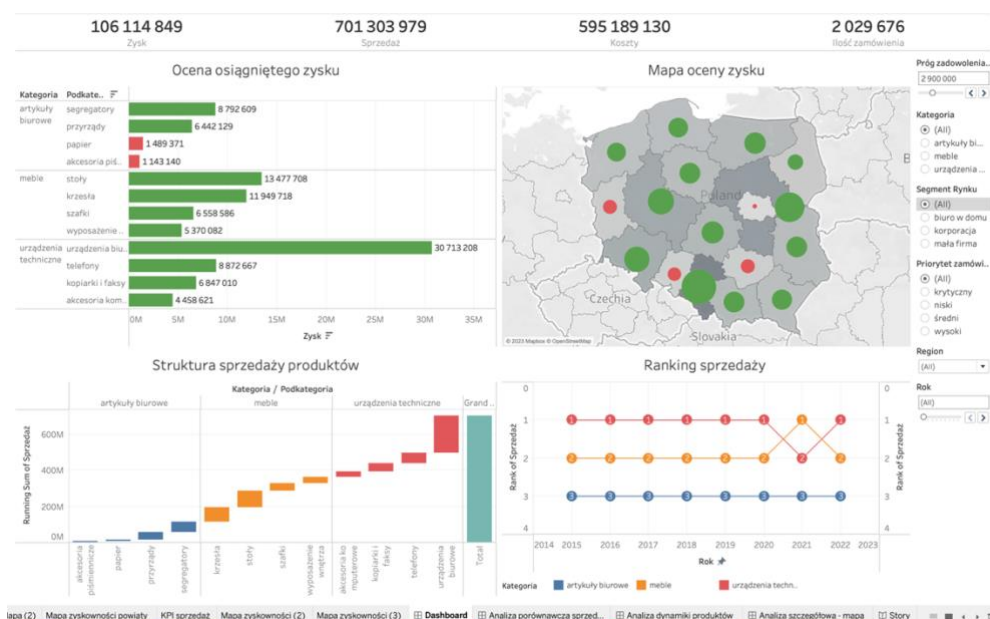
Pulpit menadżerski (ang. dashboard) to narzędzie do prezentacji danych, które pomaga menadżerom w podejmowaniu decyzji. Składa się z serii wskaźników, wykresów i tabel, które przedstawiają ważne informacje dotyczące działania firmy. Celem pulpitu menadżerskiego jest przedstawienie danych w przystępny sposób, tak aby odbiorcy mogli szybko zrozumieć sytuację oraz podjąć odpowiednie decyzje.

Modelowanie pulpitu menadżerskiego polega na stworzeniu odpowiedniej struktury oraz doborze wskaźników i wizualizacji, które pomogą odbiorcom lepiej zrozumieć sytuację firmy.

Zamodelowany prototyp pulpitu znajduje się w pliku twbx pod nazwą „projekt_260493.twbx”

- tego typu pulpit menadżerski jest interaktywny, dzięki czemu odbiorcy mogą przeglądać dane i analizować je na różnych poziomach szczegółowości, np. po dacie, po produktach, po regionie itp.

GŁÓWNY PULPIT

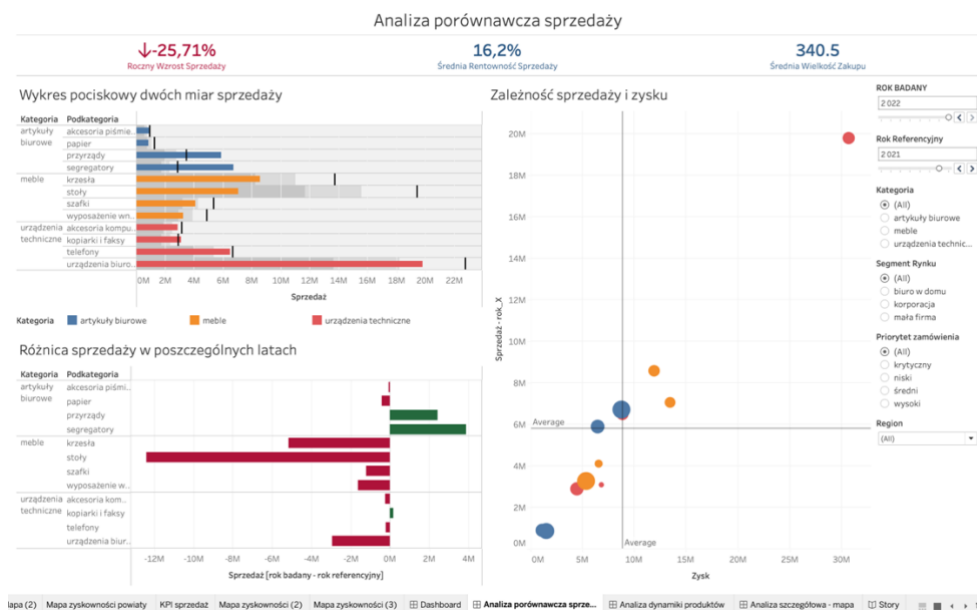


Pulpit prezentuje wartości finansowe na samej górze. Głównym zadaniem pulpitu jest określenie progu zadowolenia z zysku, a następnie analizowanie, które produkty osiągają ten poziom i które nie. Dzięki temu, możliwe jest identyfikowanie obszarów, które wymagają poprawy, oraz tych, które generują dobry zysk. Na dolnych wykresach widać strukturę sprzedaży produktów, przedstawioną w formie tzw. wykresu "waterfall", który pokazuje również skumulowaną sprzedaż poszczególnych kategorii. Dodatkowo, w prawym dolnym rogu znajduje się wykres rankingowy, który prezentuje najlepiej sprzedające się kategorie produktów w poszczególnych latach, pozwalający łatwo zauważyć trendy i tendencje na rynku. Filtry dostępne po prawej stronie pozwalają na filtrowanie danych według interesujących nas kryteriów, takich jak kategorie produktów, segment rynku, priorytety

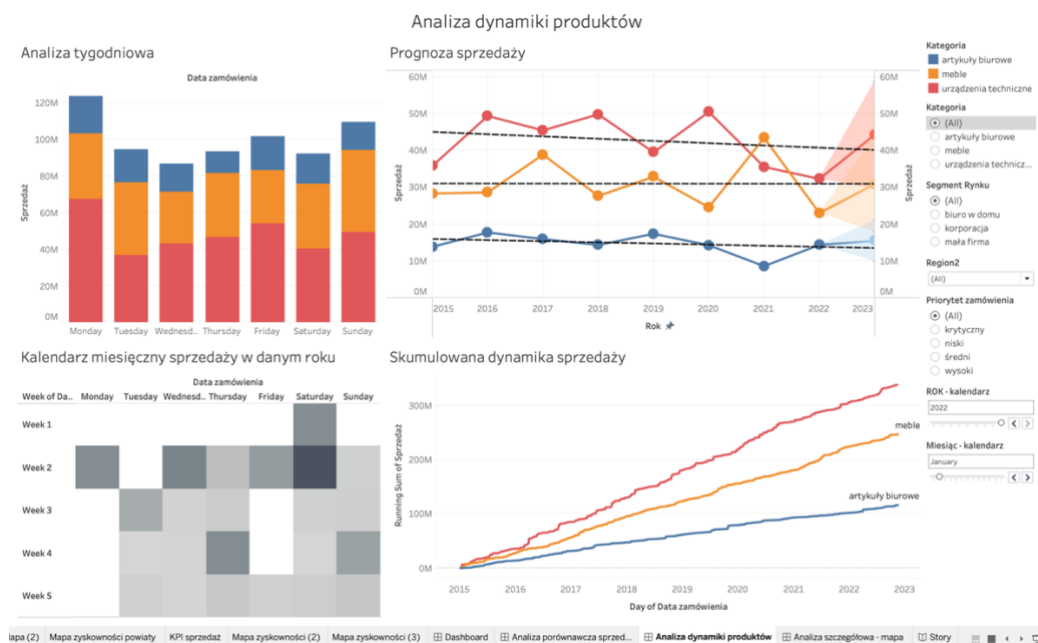
zamówień, region i rok. Pulpit ta dostarcza również narzędzia do podejmowania bardziej świadomych i informowanych decyzji biznesowych. Dodatkowo pulpit został wyposażony w możliwość drążenia danych w menu poszczególnych wykresów:

- z wykresu mapa ocen zysku jest możliwość przejścia do szczegółowej analizy związanych z danymi terytorialnymi.
- Z wykresów ocena osiągniętego zysku, struktura sprzedaży produktów i ranking sprzedaży jest możliwość przejścia do analizy porównawczej sprzedaży oraz analizy dynamiki sprzedaży.

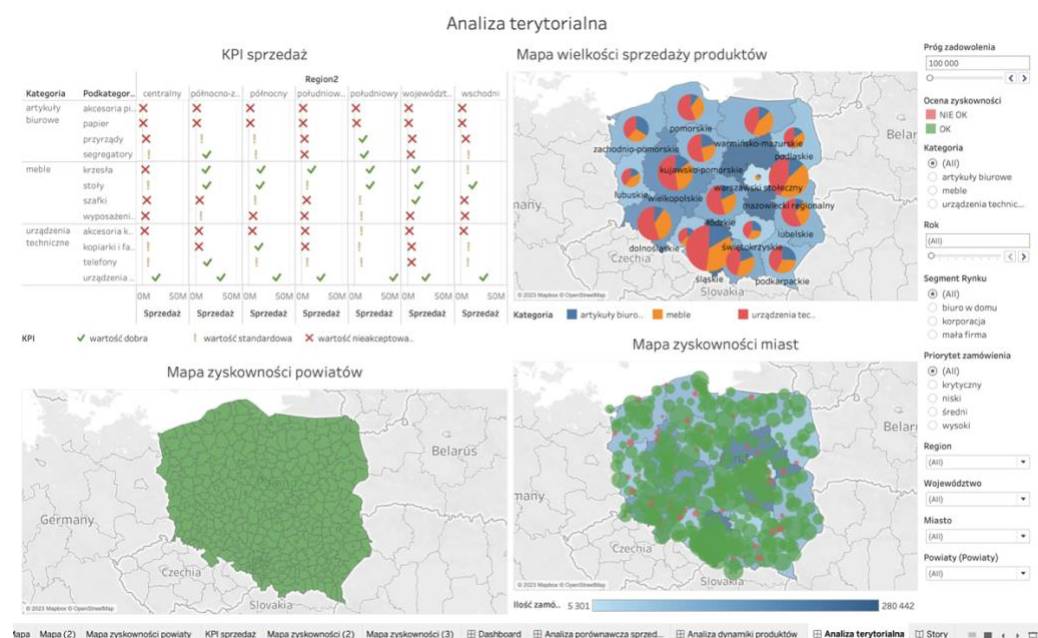
DRAŻENIE DANYCH



Analiza porównawcza sprzedaży przedstawia wykres pociskowy dwóch miar sprzedaży z badanych lat, wykres zależności sprzedaży i zysku oraz wykres różnicy sprzedaży w poszczególnych latach (gdy sprzedaż w roku badanym jest mniejsza niż w roku referencyjnym to jest to wartość zła, a gdy większa to dobra). Dane wykorzystują wskaźnik - Roczny Wzrost Sprzedaży (zdefiniowany w projekcie pulpitu)



Analiza dynamiki produktów pokazuje 4 wykresy. Analizie tygodniową, na którą składa się sprzedaż poszczególnych kategorii produktów w określone dni tygodnia. Kalendarz miesięczny sprzedaży w danym roku przedstawia sprzedaż w określonym miesiącu i roku (w przypadku zdefiniowania filtrów dla wszystkich lat i miesięcy wyniki będą nie miarodajne). Prognoza sprzedaży ukazuje dynamikę sprzedaży, która jest rozszerzona o prognozę sprzedaży na kolejny rok wraz z widełkami oraz dodatkowo zdefiniowane zostały linie referencyjne dla każdej kategorii produktu. Skumulowana dynamika sprzedaży pokazuje sumę sprzedaży we wszystkich badanych latach.



Analiza terytorialna ukazuje KPI sprzedaż (Wskaźnik efektywności sprzedaży podkategorii produktów w regionach). Mapa wielkości sprzedaży produktów jest wykresem dual axis ilości zamówień a wielkości sprzedaży konkretnych kategorii produktów. Mapa zyskowności powiatów i miast odnosi się do wcześniej wspomnianego progu zyskowności tylko na niższych poziomach NUTS.

SPRAWOZDANIE CZĘŚĆ DLA DECYDENTA – P4

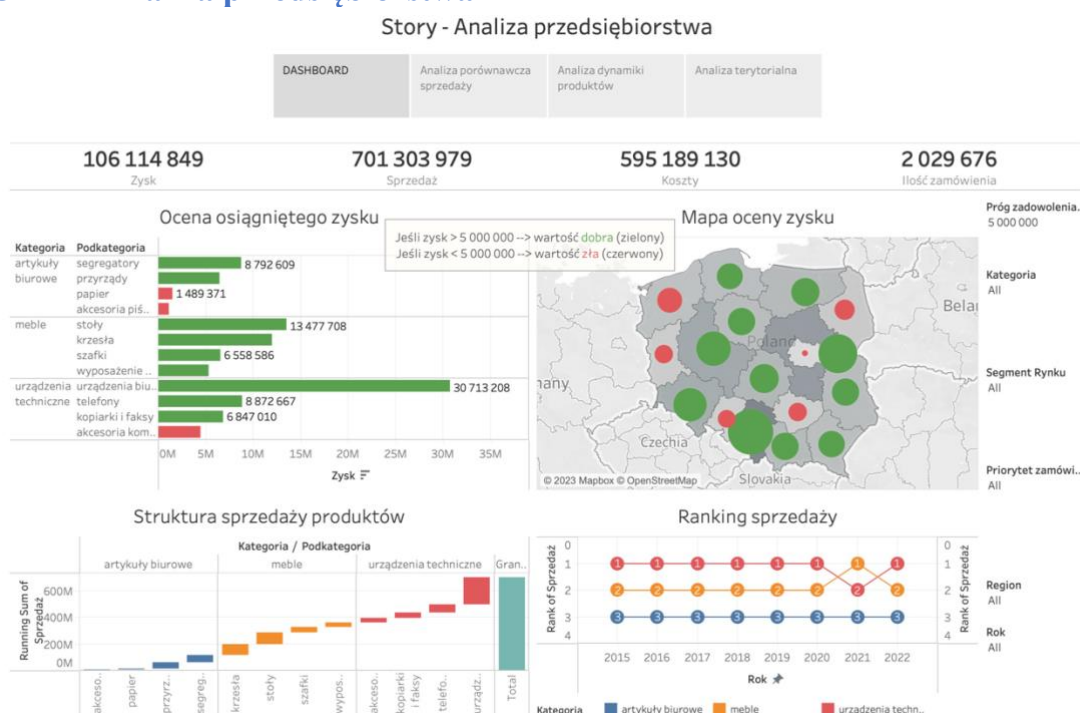
Ewaluacja i wdrożenie

Ewaluacja i wdrożenie to dwa kluczowe etapy procesu analitycznego, które pozwalają na zamianę danych w użyteczną wiedzę i działanie.

Ewaluacja to proces oceny i porównywania różnych opcji i rozwiązań na podstawie dostępnych danych. W tym etapie, analitycy przeglądają i oceniają dostępne modele, algorytmy i metody, aby wybrać najlepsze rozwiązanie dla konkretnego problemu lub celu. Ewaluacja jest ważna, ponieważ pozwala na określenie skuteczności różnych rozwiązań i zwiększenie prawdopodobieństwa wyboru najlepszego rozwiązania.

Wdrożenie to proces implementacji

STORY – Analiza przedsiębiorstwa

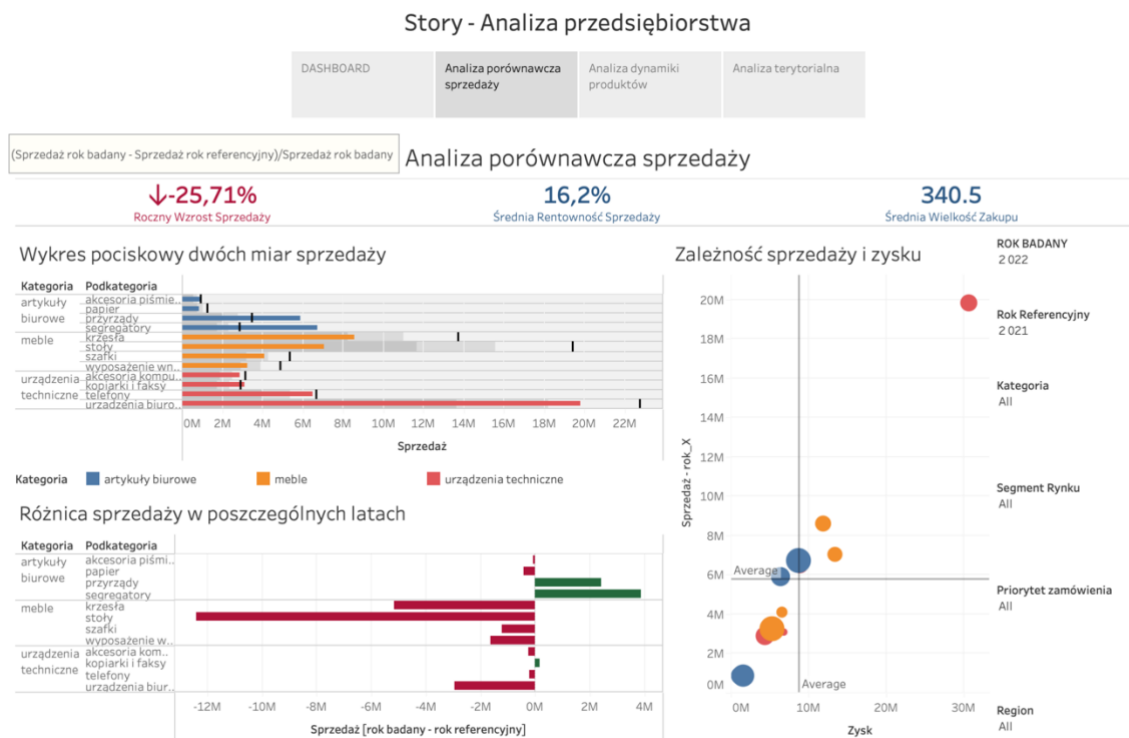


Głównym celem pulpitu jest określenie poziomu zadowolenia z osiągniętego zysku oraz identyfikowanie produktów, które osiągają ten poziom, a także tych, które nie. Dzięki temu, możliwe jest zidentyfikowanie obszarów, które wymagają poprawy oraz tych, które generują dobry zysk. Pulpit prezentuje wartości finansowe na górnej części ekranu. Na dolnej części pulpitu znajdują się wykresy, które prezentują strukturę sprzedaży produktów oraz skumulowaną sprzedaż poszczególnych kategorii. Dodatkowo, w prawym dolnym rogu znajduje się wykres rankingowy, który pokazuje najlepiej sprzedające się kategorie produktów w poszczególnych latach, co pozwala na łatwe zauważenie trendów i tendencji na rynku. Dostępne filtry po prawej stronie pozwalają na filtrowanie danych według interesujących nas kryteriów, takich jak kategorie produktów, segment rynku, priorytety zamówień, region i rok. Pulpit ten dostarcza również narzędzia do podejmowania bardziej świadomych i informowanych decyzji biznesowych.

Dodatkowo pulpit został wyposażony w możliwość drążenia danych z menu poszczególnych wykresów:

- Z wykresu mapy ocen zysku jest możliwość przejścia do szczegółowej analizy związanej z danymi terytorialnymi.
- Z wykresów oceny osiągniętego zysku, struktury sprzedaży produktów i rankingu sprzedaży jest możliwość przejścia do analizy porównawczej sprzedaży oraz analizy dynamiki sprzedaży.

Te funkcjonalności pozwalają na bardziej szczegółową analizę i wyciągnięcie więcej informacji z danych,



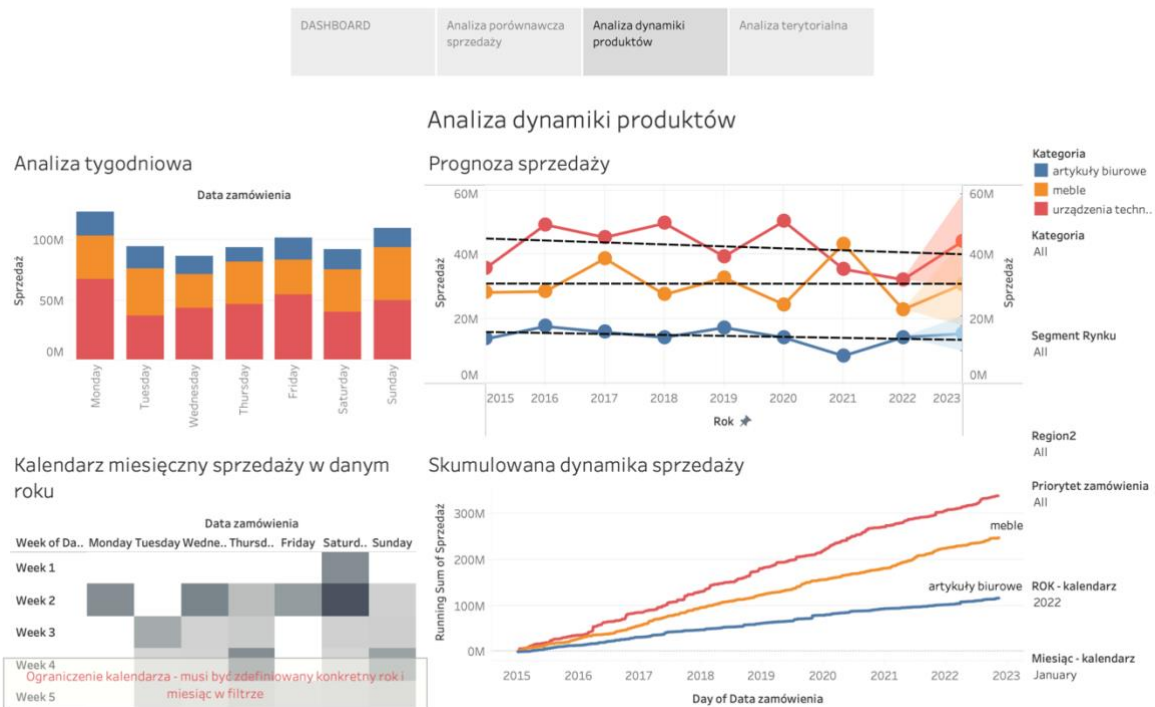
Analiza porównawcza sprzedaży polega na przedstawieniu wykresów, które ukazują zmiany sprzedaży w dwóch badanych latach. Dodatkowo zostaje przedstawiony wykres zależności sprzedaży i zysku, na którym przedstawione są cztery kategorie rozmieszczenia podkategorii produktów. Te kategorie to:

- produkty, które osiągnęły duży zysk i sprzedaż (prawy górny narożnik wykresu)
- produkty z małym zyskiem przy dużej sprzedaży (lewy górny narożnik wykresu)
- produkty z małą sprzedażą i zyskiem (lewy dolny narożnik wykresu)
- produkty z dużą sprzedażą, ale małym zyskiem (prawy dolny narożnik wykresu)

Analiza ta wykorzystuje wskaźnik - Roczny Wzrost Sprzedaży $[(\text{sprzedaż roku badanego} - \text{sprzedaż roku referencyjnego}) / \text{sprzedaż roku badanego}]$. Ten wskaźnik pozwala na określenie, czy sprzedaż wzrosła czy też spadła w porównaniu do roku referencyjnego. Dzięki temu wskaźnikowi, możliwe jest również określenie tempa wzrostu lub spadku sprzedaży.

Analiza porównawcza sprzedaży pozwala na szybkie i łatwe zidentyfikowanie produktów, które przynoszą największy zysk lub sprzedaż, a także tych, które nie przynoszą oczekiwanego zysku lub mają niski poziom sprzedaży. Dzięki temu, możliwe jest podejmowanie odpowiednich działań, takich jak zwiększanie sprzedaży produktów o niskiej sprzedaży lub redukcja kosztów produktów o niskim zysku.

Story - Analiza przedsiębiorstwa

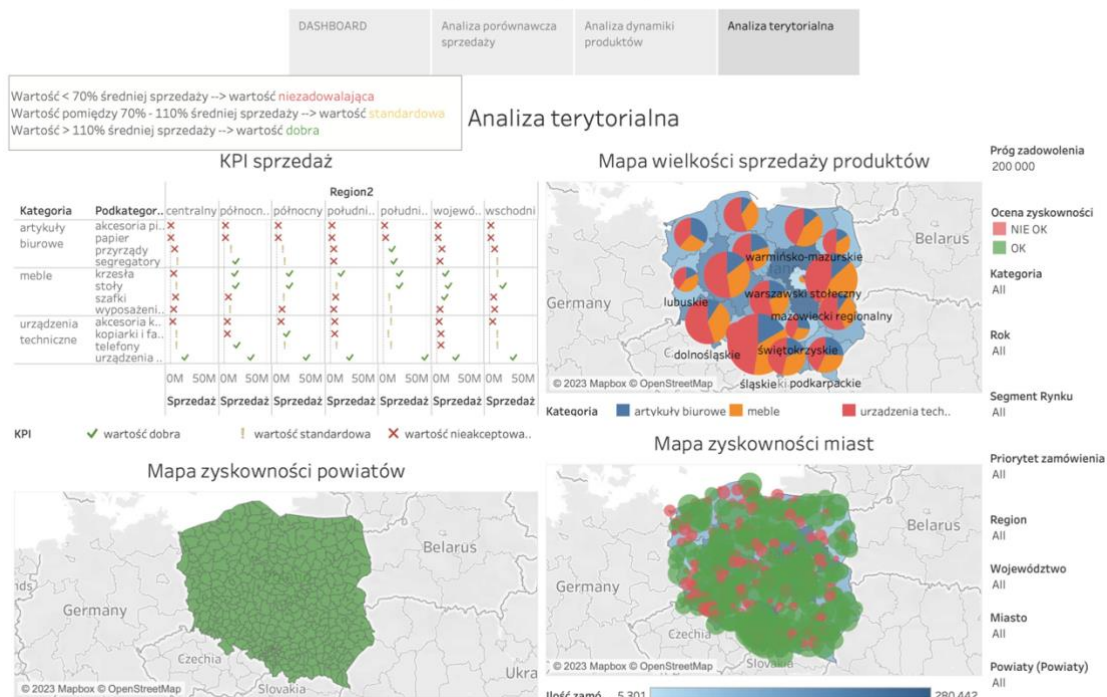


Analiza dynamiki produktów polega na przedstawieniu czterech wykresów, które pomagają zrozumieć zmiany w sprzedaży poszczególnych kategorii produktów na przestrzeni czasu. Te wykresy to:

- Analiza tygodniowa, na którą składa się sprzedaż poszczególnych kategorii produktów w określonych dniach tygodnia.
- Kalendarz miesięczny sprzedaży, który przedstawia sprzedaż w wybranym miesiącu i roku.
- Prognoza sprzedaży, która ukazuje dynamikę sprzedaży, rozszerzoną o prognozę sprzedaży na kolejny rok, z widełkami oraz zdefiniowanymi liniami referencyjnymi dla każdej kategorii produktu.
- Skumulowana dynamika sprzedaży, która pokazuje sumę sprzedaży wszystkich badanych lat.

Te wykresy pozwalają na łatwe zauważenie trendów i tendencji sprzedaży poszczególnych kategorii produktów, a także pozwalają na prognozowanie przyszłej sprzedaży i podejmowanie odpowiednich działań. Dodatkowo, dzięki analizie tygodniowej i miesięcznej, możliwe jest zidentyfikowanie dni tygodnia lub miesięcy, kiedy sprzedaż jest największa lub najmniejsza, co pozwala na dostosowanie strategii sprzedaży do tych danych.

Story - Analiza przedsiębiorstwa



Analiza terytorialna polega na przedstawieniu kluczowych wskaźników efektywności sprzedaży (KPI) podkategorii produktów w różnych regionach geograficznych. W celu przedstawienia tych danych zastosowano następujące wykresy:

- Mapa wielkości sprzedaży produktów jest to wykres, który ilustruje ilość zamówień oraz wielkość sprzedaży konkretnych kategorii produktów w poszczególnych regionach geograficznych. Może to pomóc w identyfikacji regionów, w których sprzedaż jest największa lub najmniejsza oraz w określeniu najlepiej sprzedających się produktów w danym regionie.
- Mapa zyskowności powiatów i miast odnosi się do progu zyskowności, który został wcześniej określony. Ta mapa pozwala na zobaczenie, które powiaty i miasta generują największy zysk, a także te, które są mniej zyskowne. Dzięki temu, możliwe jest identyfikowanie obszarów, które wymagają poprawy, oraz tych, które generują dobry zysk. Ta mapa jest ważnym narzędziem do podejmowania bardziej świadomych i informowanych decyzji biznesowych.

Podsumowanie

Podsumowując, analiza przedsiębiorstwa polega na przedstawieniu szeregu wykresów i map, które umożliwiają zrozumienie wyników finansowych i sprzedaży firmy. Wśród tych analiz są analiza porównawcza sprzedaży, analiza dynamiki produktów, analiza terytorialna, które pozwalają na zidentyfikowanie obszarów, które wymagają poprawy oraz tych, które generują dobry zysk. Dzięki temu, przedsiębiorstwo może podejmować bardziej świadome i informowane decyzje biznesowe.