

Indeks	Punkty	Uwagi - sprawozdanie	Punkty (prezenta	Uwagi - prezentacja
258478	4	<p>"</p> <ul style="list-style-type: none"> - za preprocessing związany z dzielnicami. - raport powinien być bardziej formalny. To nie jest opowiadanie „Model jest bardzo słaby” + diagnostyka modelu + transformacja Box-Coxa - wyrzucanie obserwacji. Czemu tak niski próg odcięcia? 2 sigma to dosyć mało. Samo wyrzucanie obserwacji jest w porządku + dodanie nowych zmiennych do modelu, namierzenie problemu - wielomian 3 stopnia na dacie zbudowania. To strasznie nieintuicyjnie, nie wiadomo jak to interpretować. Lepiej zrobić dyskretyzację, przedwojenne, wielka płyta, nowe budownictwo + wychwycenie zależności dla powierzchni - obrazki, dopasowanie są dla krzywych znacznie wyższego stopnia! - skoro nie ma jednorodnej wariancji to model nie jest poprawnie zbudowany! <p>"</p>	1	<p>"</p> <p>Brak jasnego zaznaczenia co chcemy zrobić i po co. Bardzo techniczne, rozpoczęcie od szczegółów. Dobra praktyka: wykresy, wizualizacje Bardzo mocno o danych, szczegóły o dzielnicach, co odrzucamy. Bardziej interesuje nas jaki jest zakres mieszkań, które możemy przewidywać.</p> <p>Sformułowania nie rodzące zaufanie „pewnie”,</p> <p>Sformułowania matematyczne:</p> <ul style="list-style-type: none"> * „wielomianowe” * „dopasowanie funkcji”, „wielomian trzeciego stopnia” * „mediana”, „kwantyl” * „wyjaśniliśmy liczbę pokoi powierzchnią” * „box-cox z lambda -1” * „piętro w potęgze 2” * „normalne reszty” * „testy jednorodności wariancji” <p>Brakuje zmiennych (parking, strzeżone).</p> <p>Pytania:</p> <ol style="list-style-type: none"> 1. Jak wpływają zmienne? 2. Jaka jest dokładność modelu? Jak to rozumieć? 3.

258554	3	<p>"</p> <ul style="list-style-type: none"> - raport powinien być pisany bezosobowo w czasie przeszłym np. „zbadano”, "usunięto" - „test na korelację" podczas gdy obliczany jest VIF. To raport techniczny - musi być diablo precyzyjny - strona 7, co jest na osi X? + za wykres ofert od daty budowy - kiedy robimy grupowanie zmiennej ciągłej to dostajemy zmienną dyskretną! + za podjęcie decyzji o ograniczeniu zakresu predykcji od lat 60-tych - model dla przedwojennych mieszkań, za dużo zmiennych w stosunku do danych! - Podsumowanie - co zostało zrobione, bez filozofii ;) np. „Ze względu na nieliniowość zbudowano osobne modele dla budynków przedwojennych i powojennych" - Więcej precyzji np. „dopasowań regresji na poziomie od 50 do 70 procent". To jest raport dla statystyka! 	<p>"</p> <p>Bardzo dobre wprowadzenie, charakterystyka danych - co mamy, czego nie mamy. Czego możemy się spodziewać.</p> <p>Główny zarzut: Nie mamy tłumaczyć procesu dojścia do modelu, tylko wytłumaczyć model.</p> <p>Wytłumaczenie usuwania danych ok. Można mniej mówić o tym czemu.</p> <p>Język matematyczny:</p> <ul style="list-style-type: none"> * R^2 * brak wpływu * niezależność błędu * „testy na jednorodność wariancji nie przeszły" * wykres diagnostyczny * „matematyczne przekształcenia" <p>Język statystyczny:</p> <ul style="list-style-type: none"> * techniczne wykresy * testy <p>Przed I Wojną Warszawa istniała. Bardziej chodzi o II Wojnę i zniszczenia.</p> <p>Końcówka mocno techniczna. Nie interesuje nas co zostało zrobione, ale jak to wpływa na możliwości modelu.</p> <p>Uwaga: Mieszanie przedwojenne są drogie, bo są w dobrych dzielnicach i w centrum!</p> <p>Dobra końcówka: na czym zależy firmie? Rozumiemy dane, możemy wam pomóc, określcie na czym Wam głównie zależy.</p> <p>3 Uwaga: Wawer ma dużą powierzchnię, ale nie tak dużą ludność</p>
--------	---	---	--

258486	4	<p>"</p> <ul style="list-style-type: none"> + struktura raportu + decyzja o podziale na dwie grupy - brak informacji o mieszkaniach, które zostały usunięte - usuwanie obserwacji w oparciu o odległość międzykwartlową, gołym okiem widać skośność danych - potrzebna transformacja a nie wyrzucanie obserwacji! - chaos w drugiej części raportu, dużo wypłujek eRowych, ale nie widać co jest celem <p>"</p>	<p>"</p> <p>Wstęp: mieszkanie - dla każdego lub luksusowe (ważne rozróżnienie)</p> <p>Uwaga: Kudos dla mamy - bardzo fajnie</p> <p>Uwaga: zamiast powiedzieć co zostało zrobione - jak doszedłeś do wyniku</p> <p>Powiedzenie jakie są istotne czynniki i co można powiedzieć - to jest ok. Przybliżenie tego co się znajduje w danych. To są naturalne rzeczy dla zarządu. Zaleta jest taka, że mamy wspólne pojęcia.</p> <p>Uwaga: bardzo fajne wytłumaczenie, czemu piętro ma wpływ tylko na stare budownictwo! Można by stworzyć jakieś kombinowane zmienne: wysokie piętro bez windy.</p> <p>Kudos za mapkę: czym są lokalizacje A,B,C,D? Brak powiązanie literki z kolorem! Przekształcenie, pogrupowanie w oparciu o dane - bardzo dobre. Ale można znacznie oszczędniej opowiadać.</p> <p>Uwaga: Powierzchnia mieszkania: nie widać specjlanie różnicy. Można zrobic dyskretyzację. Małe kawalerki i pozostałe.</p> <p>Język matematyczny:</p> <ul style="list-style-type: none"> * histogram * kwantyle * odstęp międzykwartylowy * R^2 <p>3 Bardzo dobre podsumowanie tego do czego sprowadza się R^2!</p>
--------	---	--	--

258536	3	<p>"</p> <ul style="list-style-type: none"> - opcja chunk message=FALSE +wyrzucenie obserwacji z poza Warszawy - regresja NIE zakłada niezależności zmiennych. Po prostu w takim wypadku trudniej budować modele - wyrzucenie zmiennych powinno być lepiej umotywowane. Może VIF? - modelujemy cenę za m², powierzchnia mieszkania może na to nie wpływać - lepiej najpierw zrobić transformację, a potem odrzucać obserwacje. Być może po transformacji obserwacja będzie mniej wpływać? - dlaczego przekształcenie Box-Coxa na zmienną surface? Powinno być na zmienną objaśnianą! - to samo dla zmiennej floor. Jako argument box-coxa można dać pełny model, nie trzeba robić modeli z jedną zmienną - obserwację dla Wawra można ewentualnie usunąć, patrząc na ceny pozostałych mieszkań w tej dzielnicy - p-wartość<0.05 to nie jest game-over. Wskazuje na potencjalny problem - brakuje podsumowania 	<p>"</p> <p>Wstęp ok: po co budujemy, z czego chcemy skorzystać, korzystamy tylko z danych liczbowych, bez wiedzy eksperckiej.</p> <p>Uwaga: Można by dodać krótkie podsumowanie danych.</p> <p>Opis metod: w porządku, przewidujemy cenę metra kwadratowego. Opis ograniczeń, czemu wyrzucamy obserwacje, sygnalizujemy, że zrobiliśmy pewne przekształcenia.</p> <p>Wykres: czego możemy oczekiwać - ok.</p> <p>Preprocessing danych: nie trzeba mówić, że pomijamy zmienne. Jak będą pytania, to musimy potrafić na nie odpowiedzieć. Nie trzeba się tłumaczyć, zanim nie zostanie zadane pytanie.</p> <p>Uwaga do wizualizacji: boxplot dobrze jest uszeregować po medianie</p> <p>Język matematyczny:</p> <ul style="list-style-type: none"> * obserwacja odstająca * wykresy techniczny normalności * ujemne wartości czego? <p>Uwaga: Efektywność predykcji, co oznacza 40%? To nie to co zostało powiedziane!</p>	3
--------	---	--	--	---

258601	3	<p>"</p> <p>+wstęp jest ok</p> <p>- rekomendacje są niechlujnie spisanie np. „Obserwacje z liczbą pokoi większą od 11, usunąć lub zaagregować do jednej.”</p> <p>- miara Cooka NIE jest duża, nie jest możliwe, żeby 78% obserwacji było powyżej punktu odcięcia</p> <p>- kilka wykresów diagnostycznych i box-coxa zupełnie bez komentarza</p> <p>- brak informacji o tym jak dobrze był dopasowany model2</p>	<p>"</p> <p>cel: bez regresja liniowa, prezentacja jest biznesowa. Schemat zależności jest dobry.</p> <p>Uwaga: kondycja mieszkania -> stan mieszkania</p> <p>Fajny wstęp: czego się dowiedzieliśmy. To są nasze założenia. Jeśli coś nie jest założeniem, to nie ma powodu o nich wspominać.</p> <p>Rozkład cen w skali logarytmicznej -> ok. Dobrze, że jest wykres.</p> <p>Zależność od liczby pokoi.</p> <p>Uwaga: na Pradze</p> <p>Język matematyczny:</p> <ul style="list-style-type: none"> * obserwacje odstające * wykres diagnostyczny <p>Uwaga: czym jest dopasowanie modelu w %?</p> <p>Uwaga: próbowałem zbadać -> zbadałem</p> <p>Uwaga: wiedza ekspercka -> ok, ale prawobrzeżna i lewobrzeżna Warszawa bardzo się różnią</p> <p>Podział na 3 okresy -> ok. Mieszkania są bardzo różne.</p> <p>3 Uwaga: Co to jest „miara predykcji”. Co i jak możemy przewidzieć?</p>
--------	---	---	--

258544	5	<p>"</p> <ul style="list-style-type: none"> + za znalezienie źródła danych - dopóki nie mamy właściwej skali trudno jest usunąć obserwacje odstające - może we właściwej skali te dane się symetryzują? - do chunka można dodać opcję warning=FALSE + bardzo fajna analiza eksploracyjna! Dobre wykresy, zrozumienie danych, włożenie wiedzy z zewnątrz, transformacja zmiennych. super! - mała uwaga do piętra: wydaje się, że 14+ powinno być osobno. To mieszkania nie w blokach, droższe + co do analizy roku budowy: można dodać zmienną kategorię i dodać jej interakcję z całą resztą - brakuje info o lambda jakie maksymalizuje likelihood w przekształceniu box-coxa. Dlaczego „Model bardzo zmodyfikowany” ma transformację logarytmiczną? Może to wynik usunięcia obserwacji odstających na początku? 		
258609	4	<p>"</p> <ul style="list-style-type: none"> - eksploracja danych. Ale część wniosków niepoprawna. Mamy np. zależność między powierzchnią a ceną metra kwadratowego - korelacje są trochę na wyrost. O ile można jeszcze przyjąć, że mieszkanie trzypokojowe jest 3/2 bardziej pokojowe niż dwupokojowe. To już trudno uznać, że 10 piętro jest 5 razy bardziej piętrowe niż drugie. To, że zmienna jest wyrażona liczbami nie oznacza, że nie jest kategorię - „performance wzrósł” - naprawdę można ładnie, precyzyjnie i po polsku + za pomysł z grupowaniem dzielnic - model z usuniętą zmienną surface. Poprawa R^2 wynika z ograniczenia zbioru do mieszkań zbudowanych po wojnie. Usunięcie zmiennej zawsze powoduje spadek R^2 + bardzo dobry pomysł ze stworzeniem nowej zmiennej - rok zbudowania budynku jako zmienna kategorię jest ryzykownym pomysłem. R^2 mocno wzrosło dlatego, że wprowadziliśmy dodatkowych 60 zmiennych do modelu. Grupowanie ma sens, ale rozważanie każdego roku osobno może doprowadzić do nadmiernego dopasowania do danych treningowych! 		

258512	4	<p>"</p> <ul style="list-style-type: none">+ połączenie dzielnic, redukcja liczby parametrów- używaj opcji chunków warning=FALSE i message=FALSE. Nie ma powodu żeby pojawiały się w raporcie+ eksploracja danych. Brakuje jedynie wniosków z histogramów- brak wniosków z przeprowadzonej diagnostyki pierwszego modelu- liczba pokoi jest w gruncie rzeczy zmienną jakościową, a nie ilościową. Mówienie o korelacji jest pewnym nadużyciem. Jeszcze bardziej nieuzasadnione jest liczenie korelacji z piętrem+ redukcja liczby parametrów przez grupowanie dzielnic. Bardzo dobre!+ „lewele”+ podział na trzy grupy ze względu na rok budowy+ usunięcie zmiennych. Porównanie przez analizę wariancji- brakuje podsumowania. Co zostało zrobione, jak wyglądają modele, jakie są ich ograniczenia		
--------	---	---	--	--

	<p>"</p> <p>- reszty mają mieć rozkład normalny. Zmienna objaśniana nie! Statystyk nie jest aptekarzem. -0.98 to -1.</p> <p>- jak liczone są korelacje? W przypadku np. piętra korelacja jest nieuzasadniona (zmienna jakościowa)</p> <p>- dodawanie zmiennej typu 3 potęga powierzchni sprawia, że model jest nieinterpretowalny. Należy tego unikać!</p> <p>- π^2 to znowu matematyczna aberracja</p> <p>3 + diagnostyka, budowa skutecznego modelu</p>	<p>"</p> <p>Uwaga: użycie plotly - bardzo fajnie. Wytłumaczenie czemu robimy przekształcenie jest ok, ale nie jest konieczne</p> <p>Opis cech - bardzo fajne. Co mamy w danych, co da się wyczytać. Wytłumaczenie dopasowania trzeciego stopnia, bardzo dobre (ewentualnej dyskretyzacja mogłaby tam być)</p> <p>Opis cech, liczba pokoi: opis zależności</p> <p>Uwaga: Dlaczego deweloperskie są takie dobre? Przecież to nie dlatego, że są w wysokim standardzie.</p> <p>Opis cech: zależność od roku budowy. „zależność trzeciego stopnia”</p> <p>Język matematyczny:</p> <ul style="list-style-type: none"> * oznacznie $f(\text{Powierzchnia}, \dots)$, szum \rightarrow zamiast tego rysunek co na wejściu, co na wyjściu * mediana \rightarrow wartość średnia * zależność kwadratowa \rightarrow po prostu można opisać <p>Współczynnik dopasowania: brak wytłumaczenie jak to wpływa.</p> <p>Uwaga: jeśli skala jest nieprzejrzysta, ale jest sensowna, to można usunąć. Najlepiej mieć</p> <p>3 skalę, która jest interpretowalna.</p>
--	--	--