

Analiza rynku mieszkaniowego z danych otodom SQL

Dokument przedstawia analizę rynku mieszkaniowego w pięciu największych miastach Polski, przeprowadzoną na podstawie danych z portalu Otodom. Celem projektu jest zidentyfikowanie czynników wpływających na cenę za metr kwadratowy oraz ocena struktury ofert pod względem metrażu, wieku budynków, stanu wykończenia i obecności udogodnień.

Analiza została wykonana w języku SQL w środowisku Google BigQuery, co umożliwiło szczegółową segmentację rynku oraz obliczenie kluczowych metryk. Wyniki pokazują m.in. wpływ windy, parkingu i tarasu na ceny, różnice cenowe między rodzajami zabudowy, a także charakterystyczne wzorce cen dla poszczególnych miast. Otrzymane wnioski mogą służyć jako podstawa dla analiz inwestycyjnych i narzędzi wspierających decyzje rynkowe.

Sprawdzenie kompletności danych oraz wpływ udogodnień i wysokości zabudowy na ceny za metr kwadratowy.

Kontrola kompletności danych dla pięciu miast pokazuje bardzo podobny wzorzec. Największe braki dotyczą certyfikatu energetycznego (ok. 87-93% NULL we wszystkich miastach) oraz pól opisowych: wyposażenie, zabezpieczenia, materiał budynku, media posiadają zwykle 40-70% braków. Informacje kluczowe dla wyceny są kompletne: *cena, metraż, cena_m², miasto, ogrzewanie, winda, taras/balkon*: 0% NULL. Atrybuty pięter są relatywnie dobrze wypełnione: *pietro_mieszkania* 2-4% braków, *najwyższe_pietro* 5-7%, *procent_wysokości* 5-9%.

Wnioski dla analiz: zmienne o bardzo niskim pokryciu (np. *certyfikat*) należy pominąć lub traktować „brak danych” jako osobną kategorię. Pola pół - strukturalne (*wyposażenie, zabezpieczenia*) warto agregować do nielicznych, binarnych cech. Metryki piętra można wykorzystywać, pamiętając o niewielkim odsetku braków.

```

STRUCT('certyfikat_energetyczny' AS kolumna, s.certyfikat_energetyczny IS NULL OR TRIM(s.certyfikat_energetyczny) = '' AS is_null),
STRUCT('informacje_dodatkowe' AS kolumna, s.informacje_dodatkowe IS NULL OR ARRAY_LENGTH(s.informacje_dodatkowe)=0 AS is_null),
STRUCT('cena_za_m2' AS kolumna, s.cena_za_m2 IS NULL AS is_null),
STRUCT('mieszkania' AS kolumna, s.mieszkania IS NULL AS is_null),
STRUCT('najwyzsze_pietro' AS kolumna, s.najwyzsze_pietro IS NULL AS is_null),
STRUCT('pietro_raw' AS kolumna, s.pietyro_raw IS NULL OR TRIM(s.pietyro_raw) = '' AS is_null),
STRUCT('pietro_normalized' AS kolumna, s.pietyro_normalized IS NULL OR TRIM(s.pietyro_normalized) = '' AS is_null),
STRUCT('procen_t_wysokosci_budynku' AS kolumna, s.procenzt_wysokosci_budynku IS NULL AS is_null),
STRUCT('garazmiesjsce_parkingowe' AS kolumna, s.garazmiesjsce_parkingowe IS NULL AS is_null),
STRUCT('piwnica' AS kolumna, s.piwnica IS NULL AS is_null)
) ) AS i
)
SELECT
miasto,
kolumna,
COUNT(*) AS liczba_wierszy,
SUM(CASE WHEN is_null THEN 1 ELSE 0 END) AS liczba_null,
COUNT(*) - SUM(CASE WHEN is_null THEN 1 ELSE 0 END) AS liczba_poprawnych,
ROUND(100 * SAFE_DIVIDE(SUM(CASE WHEN is_null THEN 1 END), COUNT(*)), 2) AS null_pct
FROM flat
GROUP BY miasto, kolumna
ORDER BY miasto, null_pct DESC, kolumna;

```

Fragment kodu analizy kompletności danych

W tej analizie sprawdzono jak trzy udogodnienia - *winda*, *miejsce parkingowe* oraz *taras/balkon* wpływają na cenę za metr kwadratowy. Podstawą obliczeń jest jednolita miara *cena_m2* (kolumna gotowa). Dla każdego miasta i segmentu stanu lokalu („do zamieszkania”, „do wykończenia”, „do remontu”, „inne”) porównywane są dwie grupy ofert: z daną cechą i bez niej. Wyznaczane są średnia i przybliżona mediana w obu grupach, a „premia” jest różnicą między nimi (*uplift*). Rekordy bez informacji o występowaniu cechy są pomijane.

```

SELECT
miasto,
stan_segment,
czynnik, -- 'winda' / 'parking' / 'taras_balkon'
COUNT(*) AS liczba_ofert_w_segmicie,
COUNTIF(has) AS n_yes,
COUNTIF(NOT has) AS n_no,

-- średnie m2
ROUND(AVG(CASE WHEN has THEN cena_m2 END), 0) AS avg_m2_yes,
ROUND(AVG(CASE WHEN NOT has THEN cena_m2 END), 0) AS avg_m2_no,
ROUND(
    AVG(CASE WHEN has THEN cena_m2 END)
    - AVG(CASE WHEN NOT has THEN cena_m2 END)
    , 0) AS uplift_avg_m2,

```

Fragment kodu analizy wpływu windy, parkingu i balkonu na cenę

Wyniki wskazują, że winda i parking zwykle podnoszą cenę o kilkaset PLN/m². Taras/balkon częściej zwiększa średnią niż medianę, co sugeruje wpływ pojedynczych, bardzo drogich ofert. Interpretacja powinna uwzględnić liczebność i rozkłady w poszczególnych segmentach.

Następny etap analizy estymuje, jak wartość windy zmienia się w zależności od piętra mieszkania. Po oczyszczeniu danych tworzone są koszyki pięter od 1 do 9 oraz grupa „10+”. W obliczeniach uwzględniono wyłącznie rekordy z podaną informacją o windzie i piętrze ≥ 1 (w tym przypadku oznacza to parter). W każdym koszyku

porównywane są dwie grupy: mieszkania z windą i bez windy. Raportowane są liczebności obu grup, średnia i mediana ceny za m² oraz premia liczona jako różnica między nimi w PLN/m² i procentowy wzrost ceny za mieszkania z windą.

```
SELECT
    miasto,
    pietro_bucket,
    COUNT(*) AS liczba_ofert,
    COUNTIF(winda_yes) AS n_winda,
    COUNTIF(NOT winda_yes) AS n_bez_windy,

    -- średnie m2 (TAK/NIE)
    ROUND(AVG(IF(winda_yes, cena_m2, NULL)), 0) AS avg_m2_winda,
    ROUND(AVG(IF(NOT winda_yes, cena_m2, NULL)), 0) AS avg_m2_bez_windy,

    -- premia (różnica TAK - NIE)
    ROUND(AVG(IF(winda_yes, cena_m2, NULL)) - AVG(IF(NOT winda_yes, cena_m2, NULL)), 0) AS premia_avg_m2,
    ROUND(100 * SAFE_DIVIDE(
        AVG(IF(winda_yes, cena_m2, NULL)) - AVG(IF(NOT winda_yes, cena_m2, NULL)),
        AVG(IF(NOT winda_yes, cena_m2, NULL))
    ), 1) AS premia_avg_pct,
```

Fragment kodu analizy wpływu windy na cenę w zależności od piętra

Wyniki dla wszystkich miast pokazują spójny wzorzec: premia rośnie od parteru do pięter średnich (3-6), gdzie zwykle wynosi +1,3-3,0 tys. PLN/m² (ok. 8-22%), a lokalnie sięga nawet +3,5 tys. PLN/m² (20-29%). Na wyższych piętrach (7-9) efekt jest słabszy i mniej stabilny ze względu na małe liczebności grup bez windy, co podnosi niepewność wyników. W koszyku „10+” obserwacje są zróżnicowane: w części rynków premia utrzymuje się, a w innych zanika.

Segmentacja rynku mieszkaniowego i czynniki kształtujące ceny.

W niniejszej analizie zastosowano podział rynku na segmenty według decylów. Decyl pierwszy (dla uporządkowanych rosnąco danych statystycznych) oznacza, że 10% jednostek ma wartości cechy mniejsze lub równe od decyla pierwszego, a 90% jednostek ma wartości cechy równe lub większe od decyla pierwszego.

Dla każdej lokalizacji wyznaczono trzy progi na rozkładzie ceny za m²: *p10* (dolny decyl), *p50* (mediana) oraz *p90* (górny decyl). Do obliczeń wykorzystano kolumnę *cena_m²*, przed obliczeniami ujednolicono nazwy miast, aby zapewnić spójność agregacji. Progi obliczono w *BigQuery* funkcją *APPROX_QUANTILES*, która zwraca granice kwantyli w trybie przybliżonym, wystarczającym dla dużych zbiorów danych. Próba rzędu 5 300 obserwacji na miasto zapewnia stabilne wyznaczenie skrajnych kwantyli (*p10/p90*) i zmniejsza wrażliwość wyników na pojedyncze wartości skrajne.

```

SELECT
    miasto,
    COUNT(*) AS liczba_ofert,
    ROUND(AVG(cena_m2),0) AS sr_m2,
    ROUND(APPROX_QUANTILES(cena_m2,101)[OFFSET(10)],0) AS p10_m2,
    ROUND(APPROX_QUANTILES(cena_m2,101)[OFFSET(50)],0) AS p50_m2,
    ROUND(APPROX_QUANTILES(cena_m2,101)[OFFSET(90)],0) AS p90_m2
FROM src
GROUP BY miasto
ORDER BY miasto;

```

Wyznaczenie progów cenowych 10%, 50% i 90%

Wyniki pokazują spodziewaną hierarchię cen: Warszawa: mediana 17 036 PLN/m² ($p10-p90$: 12 870-24 719), Kraków: 15 500 (12 000-21 000), Gdańsk: 13 930 (10 400-23 477), Wrocław: 12 991 (10 110-17 800), Poznań: 12 053 (9 284-16 111). Szerokość przedziału $p90-p10$ jest największa w Gdańsku (13 077) i Warszawie (11 849), a najmniejsza w Poznaniu (6 827), co wskazuje na relatywnie bardziej jednorodny rynek poznański.

Tak zdefiniowane progi pozwalają zbudować praktyczną segmentację: poziom „budżetowy” ($\leq p10$), „rynkowy” (około mediany) i „premium” ($\geq p90$), przydatną w wizualizacjach i porównaniach między miastami.

Następne badanie wiąże typ ogrzewania stosowanego w mieszkaniach z poziomem cen m², porównując mediany w pięciu miastach.

```

SELECT
    miasto, ogrzewanie,
    COUNT(*) AS n,
    ROUND(AVG(cena_m2),0) AS sr_m2,
    ROUND(APPROX_QUANTILES(cena_m2,101)[OFFSET(50)],0) AS mediana_m2
FROM src
WHERE ogrzewanie IS NOT NULL AND ogrzewanie <> ''
GROUP BY miasto, ogrzewanie
HAVING n >= 10
ORDER BY miasto, sr_m2 DESC;

```

Fragment kodu analizy zależności typów ogrzewania a ceny za metr kwadratowy.

Najczęściej najwyżej wyceniane są mieszkania z ogrzewaniem miejskim (lub w grupie „inne”, która bywa nośnikiem nowych, efektywnych systemów), co koresponduje z nowszym standardem budynków i centralnymi lokalizacjami. Z kolei ogrzewanie elektryczne w czterech z pięciu miast okazuje się najtańsze w medianie — wyjątkiem jest Kraków, gdzie mediana dla elektrycznego jest najwyższa. Ten odmienny wynik należy interpretować ostrożnie ze względu na małe próby w tej kategorii (np. Warszawa n=17, Poznań n=56), które zwiększały wrażliwość metryk na specyfikę niszowych ofert (małe, nowe lub premium). Gazowe oraz kotłownia zwykle lokują się pomiędzy elektrycznym a

miejskim. Klasa „brak informacji” przeważnie zbliża się do median rynkowych, ale utrudnia precyzyjną interpretację. Ogólny wniosek: rynek premiuje systemy postrzegane jako wygodne i przewidywalne kosztowo (miejskie), a dyskontuje warianty o potencjalnie wyższych rachunkach (elektryczne) lub starszej infrastrukturze.

Wiersz	miasto	ogrzewanie	n	sr_m2	mediana_m2
1	Gdańsk	miejskie	3138	16388.0	14281.0
2	Gdańsk	inne	91	16031.0	15905.0
3	Gdańsk	brak informacji	1597	15200.0	13550.0
4	Gdańsk	kotłownia	83	14526.0	12684.0
5	Gdańsk	elektryczne	58	13734.0	12449.0
6	Gdańsk	gazowe	332	13712.0	12500.0
7	Kraków	elektryczne	96	17496.0	17072.0
8	Kraków	miejskie	2981	16680.0	15736.0

Raport zależności typów ogrzewania od ceny za metr kwadratowy

Jednym z kluczowych czynników wpływających na wartość nieruchomości jest wiek budynku, wobec czego zasób ofert podzielono na koszyki roczników budowy, policzono ich udział procentowy na w bazie danych, średnie ceny m² oraz średni metraż w określonych przedziałach.

```

ROUND(100 * SAFE_DIVIDE(SUM(CASE WHEN rok_budowy BETWEEN 2000 AND 2020 THEN 1 ELSE 0 END), COUNT(*)), 1) AS pct_2000_2020,
ROUND(100 * SAFE_DIVIDE(SUM(CASE WHEN rok_budowy > 2020 THEN 1 ELSE 0 END), COUNT(*)), 1) AS pct_gt_2020,
ROUND(100 * SAFE_DIVIDE(SUM(CASE WHEN rok_budowy IS NULL THEN 1 ELSE 0 END), COUNT(*)), 1) AS pct_brak_danych,

-- Średnia cena za m2 w koszykach
ROUND(AVG(CASE WHEN rok_budowy < 1940 THEN cena_m2 END), 0) AS avg_m2_lt_1940,
ROUND(AVG(CASE WHEN rok_budowy BETWEEN 1940 AND 1960 THEN cena_m2 END), 0) AS avg_m2_1940_1960,
ROUND(AVG(CASE WHEN rok_budowy BETWEEN 1960 AND 1980 THEN cena_m2 END), 0) AS avg_m2_1960_1980,
ROUND(AVG(CASE WHEN rok_budowy BETWEEN 1980 AND 2000 THEN cena_m2 END), 0) AS avg_m2_1980_2000,
ROUND(AVG(CASE WHEN rok_budowy BETWEEN 2000 AND 2020 THEN cena_m2 END), 0) AS avg_m2_2000_2020,
ROUND(AVG(CASE WHEN rok_budowy > 2020 THEN cena_m2 END), 0) AS avg_m2_gt_2020,
ROUND(AVG(CASE WHEN rok_budowy IS NULL THEN cena_m2 END), 0) AS avg_m2_brak_danych,

-- średni metraż (m2) w koszykach
ROUND(AVG(CASE WHEN rok_budowy < 1940 THEN metraz_m2 END), 1) AS avg_metraz_lt_1940,
ROUND(AVG(CASE WHEN rok_budowy BETWEEN 1940 AND 1960 THEN metraz_m2 END), 1) AS avg_metraz_1940_1960,

```

Obliczanie udziału rynkowego, średnich wartości cen i metrażu w różnych przedziałach roczników budowy

W Gdańsku, Krakowie i Poznaniu ponad połowa ogłoszeń dotyczy budynków po 2020 r. (ok. 50-52%), podczas gdy w Warszawie dominują lata 2000-2020 (29%) i wciąż znaczący jest starszy zasób. Ważne zastrzeżenie: „nowsze” nie zawsze znaczy droższe. W czterech z pięciu miast średnia za m² dla rocznika 2000-2020 jest równa lub wyższa niż dla >2020 (np. Warszawa: 19,6 tys. i 18,8 tys. PLN/m²; Gdańsk: 17,8 tys. i 15,8 tys.), co sugeruje silny wpływ lokalizacji i metrażu (nowsze inwestycje częściej na peryferiach i o mniejszych powierzchniach). Dla rzetelnej interpretacji rekomendowane są porównania wewnętrz tych samych lokalizacji (np. w siatce kafli) lub modele hedoniczne, które

oddzielają efekt miejsca od wieku budynku. Dodatkowo należy uwzględnić brakujące lata budowy (np. Poznań: 20%), aby nie przeceniać różnic pomiędzy kohortami. Wyniki traktować jako trend kierunkowy, a nie miarę „premii za nowość” w całym mieście.

w Poznaniu, Gdańsku, Wrocławiu i Krakowie największe średnie powierzchnie mają mieszkania zbudowane przed 1940 r. (od ok. 67,5 m² w Poznaniu do 70,5 m² w Gdańsku, dla Warszawy ≈ 64,7 m²), co sugeruje, że najstarsze zasoby to często przestronne kamienice. W okresie 1940-1960 metraże spadają osiągając średnie wartości rzędu ok. 47-57 m² co odpowiada powojennej zabudowie z małymi, funkcjonalnymi układami.

Pomiary z bazy danych wskazują, że w latach 1960-1980 średnie metraże mieszkań spadły do ok. 47-50 m². W kolejnym przedziale 1980-2000 obserwacje wskazują wzrost średnich metraży do przedziału ok. 53-68 m² przy czym Warszawa wyróżnia się rekordową średnią na poziomie 68.1 m². W latach 2000-2020 średnie wyniki są bardzo zbliżone do poprzedniego przedziału i w większości miast różnica sięga maksymalnie poziomu 2%, wyjątkiem jest Wrocław gdzie możemy zaobserwować wzrost średniego metrażu o ≈ 9.5%. Dla mieszkań wybudowanych po 2020 roku odnotowano kolejny spadek a metraże mieszczą się w przedziale ok 52-60 m².

pct_gt_2020	pct...	avg...	avg_m2_1940_19...	avg_m2_1940_19...	avg_m2_1980_20...	avg_m2_2000_20...	avg_m2_gt_2020	avg...	avg_metr...
50.2	8.9	142...	15919.0	14164.0	12809.0	17756.0	15815.0	152...	70.5
52.1	8.1	194...	15043.0	14389.0	13851.0	16807.0	16543.0	159...	68.9
50.8	20.0	113...	10534.0	10805.0	10561.0	13272.0	13361.0	115...	67.5
26.8	8.3	228...	19239.0	16280.0	15071.0	19603.0	18817.0	177...	64.7
34.0	14.4	116...	12342.0	12715.0	11971.0	14864.0	14811.0	128...	69.9

Raport zależności cen i metrażu wobec mieszkań wybudowanych w danych przedziałach lat

W kolejnej analizie zwrócono uwagę na zależność cen za metr kwadratowy od ich wielkości. Mieszkania zgrupowano w przedziałach: poniżej 25 m², 25-40 m², 40-60 m², 60-80 m² oraz powyżej 80 m². Struktura popytu jest podobna we wszystkich miastach: dominuje segment 40-60 m² (≈ 39-44% ofert), dalej 25-40 m² (≈ 20-23%) i 60-80 m² (≈ 22-26%). Skrajne koszyki są rzadsze: <25 m² to ok. 1-2% ogłoszeń, a ≥80 m² - 10-15%.

Cenowo widać klasyczny efekt skali: najmniejsze lokale mają najwyższą cenę za m², potem stawki spadają wraz z metrażem. W części rynków pojawia się jednak **odbicie dla ≥80 m²** (np. Warszawa 20 893 vs 17 639 PLN/m² dla 60-80; Gdańsk 17 306 vs 15 085; Kraków 17 091 vs 15 146), co sugeruje lepsze lokalizacje i standard wykończenia dużych mieszkań.

Wrocław nie pokazuje wyraźnej „premii” dla największych metraży. Poznań jest najtańszy w przekrojach, Warszawa - najdroższa w każdym koszyku.

Wiersz //	miasto //	liczba_ofert //	pct_lt_25 //	pct_25_40 //	pct_40_60 //	pct_60_80 //	pct_ge_80 //	sr_m2_lt_25 //	sr_m2_25_40 //
1	Gdańsk	5300	0.8	21.3	39.6	25.6	12.6	21691.0	16053.0
2	Kraków	5300	2.2	20.6	41.2	21.7	14.2	17936.0	17584.0
3	Poznań	5300	1.5	23.2	43.6	21.7	10.0	15604.0	13492.0
4	Warszawa	5300	1.8	21.0	40.7	21.9	14.6	21523.0	18889.0
5	Wrocław	5300	2.0	21.3	38.8	24.9	13.0	16675.0	15304.0

Udział metraży i zależność ceny od wielkości

Przegląd rynku względem stanu wykończenia klasyfikuje ogłoszenia do trzech segmentów: do zamieszkania, do wykończenia oraz do remontu przy czym pozostałe trafiają do kategorii „inne”. Pozwala to określić, jaka część oferty w danym mieście jest gotowa do natychmiastowego użytkowania, a jaka wymaga dodatkowych nakładów inwestycyjnych. Dla każdego segmentu obliczono również średnią cenę za metr kwadratowy oraz ich procentowy udział w rynku dla danego miasta co umożliwia porównanie, jak stan techniczny wpływa na poziomy cen.

```
CASE
    WHEN REGEXP_CONTAINS(LOWER(COALESCE(stan_wykonczania, '')),
        r'pod\s*klucz|po\s*remontie|do\s*zamieszkania|wyko(n|ń)czon|gotow') THEN 'do_zamieszkania'
    WHEN REGEXP_CONTAINS(LOWER(COALESCE(stan_wykonczania, '')),
        r'do\s*remontu|do\s*odświe|od\s*świe') THEN 'do_remontu'
    WHEN REGEXP_CONTAINS(LOWER(COALESCE(stan_wykonczania, '')),
        r'deweloperski|do\s*wyko(n|ń)czenia|stan\s*surowy') THEN 'do_wykonczania'
    ELSE 'inne'
END AS stan_segment,
COALESCE(cena_za_m2, SAFE_DIVIDE(cena, metraz_m2)) AS cena_m2
FROM `realstate-market-poland.prod.offers_sample_per_city`
WHERE LOWER(miasto) IN ('poznan', 'poznań', 'kraków', 'krakow', 'gdańsk', 'gdansk', 'warszawa', 'wrocław', 'wroclaw')
    AND (cena_za_m2 IS NOT NULL OR (cena IS NOT NULL AND metraz_m2 IS NOT NULL AND metraz_m2 > 0))
),
```

Klasyfikacja stanu wykończenia mieszkań w danych rynkowych

W pięciu miastach dominuje segment „do zamieszkania” (największy udział ofert) z najwyższą ceną za m²—np. Warszawa ≈ 19,4 tys. PLN/m², Kraków ≈ 17,2 tys., Gdańsk ≈ 17,0 tys. Tuż za nim plasuje się „do wykończenia” (zwykle druga co do liczby grupa) ze średnimi stawkami—np. Warszawa ≈ 17,3 tys., Wrocław ≈ 13,8 tys., Poznań ≈ 13,1 tys. Najmniej liczne są ogłoszenia „do remontu” i to one najczęściej mają najniższe ceny ok. 9,5-15,3 tys. PLN/m² w zależności od miasta. Kategoria „inne” bywa niejednoznaczna i w części rynków cenowo zbliża się do „do zamieszkania”, co sugeruje różnorodny standard opisów. W ujęciu przekrojowym Warszawa pozostaje najdroższa, Poznań jest najtańszy, a Kraków i Gdańsk utrzymują poziomy pośrednio-wysokie. Wnioski: stan wykończenia silnie różnicuje cenę, a rozkłady udziałów są spójne między miastami.

Wiersz	miasto	stan_segment	liczba_ofert	udzial_pct	sr_cena_m2
6	Kraków	do_wykonczenia	2226	42.0	16074.0
7	Kraków	do_zamieszkania	2158	40.7	17193.0
8	Kraków	inne	601	11.3	15633.0
9	Poznań	do_remontu	289	5.5	9451.0
10	Poznań	do_wykonczenia	2460	46.4	13121.0
11	Poznań	do_zamieszkania	1806	34.1	12488.0
12	Poznań	inne	745	14.1	11825.0
13	Warszawa	do_remontu	471	8.9	15319.0
14	Warszawa	do_wykonczenia	1139	21.5	17347.0
15	Warszawa	do_zamieszkania	2983	56.3	19398.0

Udział stanów wykończenia i poziom cen w miastach

Zapytanie grupuje oferty według miast i rodzaju materiału z którego wykonany jest budynek, zachowując jedynie te kombinacje, które mają co najmniej 10 obserwacji, aby ograniczyć wpływ małych prób na wyniki. Dla każdej grupy obliczana jest liczebność ofert oraz średnia cena za metr kwadratowy co pozwala porównać, jak materiał konstrukcyjny wpływa na poziom cen w różnych miastach.

```

    LOWER(TRIM(material_budynku)) AS material,
    COALESCE(cena_za_m2, SAFE_DIVIDE(cena, metraz_m2)) AS cena_m2
    FROM `realstate-market-poland.prod.offers_sample_per_city`
    WHERE (cena_za_m2 IS NOT NULL OR (cena IS NOT NULL AND metraz_m2 > 0))
)
SELECT
    miasto, material,
    COUNT(*) AS n,
    ROUND(AVG(cena_m2),0) AS sr_m2
FROM src
WHERE material IS NOT NULL
GROUP BY miasto, material
HAVING n >= 10
ORDER BY miasto, material;

```

Fragment kodu analizy średniej ceny za m² w zależności od materiału budynku

Wszystkie miasta tworzą spójny wzorzec cenowy powiązany z materiałem. Najdroższe są mieszkania w konstrukcjach żelbetowych/betonowych, co zwykle oznacza nowsze inwestycje np. Gdańsk żelbet ≈ 23,7 tys. PLN/m², Warszawa beton ≈ 21,9 tys., Kraków żelbet ≈ 18,0 tys. Środek stawki zajmują cegła i pustak (zwykle 14-18 tys. PLN/m² w zależności od miasta). Najniższe poziomy notuje „wielka płyta” — wyraźny dyskonto względem segmentu premium (np. Warszawa ≈ 15,3 tys., Poznań ≈ 10,6 tys., Wrocław ≈ 12,3 tys. PLN/m²). Kategoria „innego” bywa niejednorodna, ale w części rynków zbliża się do poziomu nowego budownictwa (np. Gdańsk ≈ 17,6 tys., Warszawa ≈ 17,9 tys.). Różnice wynikają nie tylko z materiału: silnie współgrają z wiekiem budynku, lokalizacją i standardem.

Wiersz	miasto	material	n	sr_m2
1	Gdańsk	beton	21	17375.0
2	Gdańsk	beton komórkowy	61	14345.0
3	Gdańsk	cegła	856	14752.0
4	Gdańsk	inny	187	17585.0
5	Gdańsk	pustak	264	14944.0
6	Gdańsk	silikat	210	13946.0
7	Gdańsk	wielka płyta	244	13888.0
8	Gdańsk	żelbet	56	23721.0
9	Kraków	beton	95	18042.0
10	Kraków	cegła	794	17034.0

Raport zależności średniej cena za m² od materiału budynku

Kolejne zapytanie po normalizacji danych z tabeli grupuje rekordy według miasta i rodzaju zabudowy, obliczając liczbę ofert oraz średnią cenę za metr kwadratowy w każdej grupie. Wynik obejmuje jedynie te grupy, które zawierają co najmniej 10 rekordów, co zwiększa wiarygodność analizy.

Struktura zabudowy silnie różnicuje ceny m². Najwyższe poziomy notują apartamentowce (np. Warszawa ≈ 21,4 tys. PLN/m², Gdańsk ≈ 21,0 tys., Kraków ≈ 19,1 tys., Wrocław ≈ 16,2 tys., Poznań ≈ 14,2 tys.), co odzwierciedla nowsze standardy i centralne lokalizacje. Kamienice są drogie na rynkach premium (Warszawa ≈ 21,4 tys., Kraków ≈ 19,6 tys.), ale w miastach z większym zróżnicowaniem stanu technicznego bywają tańsze (Wrocław ≈ 11,9 tys., Poznań ≈ 13,1 tys.). Bloki tworzą stabilny „środek” (ok. 12-16,5 tys. w zależności od miasta). Najniższe średnie pojawiają się przy szeregowkach i domach wolnostojących (często peryferyjne lokalizacje i większe metraże obniżają ceny za m²; np. Poznań ≈ 9,4-9,7 tys.)

Wiersz	miasto	rodzaj	n	sr_m2
17	Warszawa	apartamentowiec	1234	21438.0
18	Warszawa	blok	2721	16485.0
19	Warszawa	dom wolnostojący	17	15892.0
20	Warszawa	kamienica	485	21436.0
21	Warszawa	szeregowiec	28	13324.0
22	Wrocław	apartamentowiec	1028	16168.0
23	Wrocław	blok	2287	13087.0
24	Wrocław	dom wolnostojący	53	12637.0
25	Wrocław	kamienica	751	11902.0
26	Wrocław	plomba	67	13391.0
27	Wrocław	szeregowiec	61	12011.0

Kody SQL

Ile więcej zapłacisz za winde/parking/taras

```
WITH base AS (
    SELECT
        CASE
            WHEN LOWER(miasto) IN ('poznan','poznań') THEN 'Poznań'
            WHEN LOWER(miasto) IN ('kraków','krakow') THEN 'Kraków'
            WHEN LOWER(miasto) IN ('gdańsk','gdansk') THEN 'Gdańsk'
            WHEN LOWER(miasto) = 'warszawa'      THEN 'Warszawa'
            WHEN LOWER(miasto) IN ('wrocław','wroclaw') THEN 'Wrocław'
            ELSE miasto
        END AS miasto,
        -- segment stanu
        CASE
            WHEN REGEXP_CONTAINS(LOWER(COALESCE(stan_wykonczenia,"")),
                r'pod\s*klucz|po\s*remoncie|do\s*zamieszkania|wyko(n|ń)czon|gotow') THEN 'do_zamieszkania'
            WHEN REGEXP_CONTAINS(LOWER(COALESCE(stan_wykonczenia,"")),
                r'do\s*remontu|do\s*odświej|od\s*świe') THEN 'do_remontu'
            WHEN REGEXP_CONTAINS(LOWER(COALESCE(stan_wykonczenia,"")),
                r'dewelopersk|do\s*wyko(n|ń)czenia|stan\s*surowy') THEN 'do_wykonczenia'
            ELSE 'inne'
        END AS stan_segment,
        COALESCE(cena_za_m2, SAFE_DIVIDE(cena, metraz_m2)) AS cena_m2,
        winda,
        garazmiejscie_parkingowe AS parking,
        taras_balkon
    FROM `realstate-market-poland.prod.offers_sample_per_city`
    WHERE (cena_za_m2 IS NOT NULL OR (cena IS NOT NULL AND metraz_m2 IS NOT NULL AND metraz_m2 > 0))
),
-- sprowadzamy 3 czynniki do formatu "long"
```

```

flags AS (
    SELECT miasto, stan_segment, cena_m2, 'winda'      AS czynnik, winda AS has
    FROM base
    UNION ALL
    SELECT miasto, stan_segment, cena_m2, 'parking'   AS czynnik, parking AS has
    FROM base
    UNION ALL
    SELECT miasto, stan_segment, cena_m2, 'taras_balkon' AS czynnik, taras_balkon AS has
    FROM base
)

```

```

SELECT
miasto,
stan_segment,
czynnik,          -- 'winda' / 'parking' / 'taras_balkon'
COUNT(*)          AS liczba_ofert_w_segmicie,
COUNTIF(has)     AS n_yes,
COUNTIF(NOT has) AS n_no,

```

```

-- średnie m2
ROUND(AVG(CASE WHEN has THEN cena_m2 END), 0) AS avg_m2_yes,
ROUND(AVG(CASE WHEN NOT has THEN cena_m2 END), 0) AS avg_m2_no,
ROUND(
    AVG(CASE WHEN has THEN cena_m2 END)
    - AVG(CASE WHEN NOT has THEN cena_m2 END)
, 0) AS uplift_avg_m2,

```

```

-- mediany m2 (aproksymowane)
ROUND(APPROX_QUANTILES(IF(has, cena_m2, NULL), 101)[OFFSET(50)], 0) AS med_m2_yes,
ROUND(APPROX_QUANTILES(IF(NOT has, cena_m2, NULL), 101)[OFFSET(50)], 0) AS med_m2_no,
ROUND(
    APPROX_QUANTILES(IF(has, cena_m2, NULL), 101)[OFFSET(50)]
    - APPROX_QUANTILES(IF(NOT has, cena_m2, NULL), 101)[OFFSET(50)]
, 0) AS uplift_med_m2

```

```

FROM flags
WHERE has IS NOT NULL          -- odrzuć rekordy bez informacji o czynniku
GROUP BY miasto, stan_segment, czynnik
ORDER BY miasto, stan_segment, czynnik;

```

Ile zapłacisz więcej za windę zależnie od piętra

```

WITH base AS (
    SELECT
        CASE
            WHEN LOWER(miasto) IN ('poznan','poznań') THEN 'Poznań'
            WHEN LOWER(miasto) IN ('kraków','krakow') THEN 'Kraków'
            WHEN LOWER(miasto) IN ('gdańsk','gdansk') THEN 'Gdańsk'
            WHEN LOWER(miasto) = 'warszawa'      THEN 'Warszawa'
            WHEN LOWER(miasto) IN ('wrocław','wroclaw') THEN 'Wrocław'
            ELSE miasto
        END AS miasto,
        COALESCE(cena_za_m2, SAFE_DIVIDE(cena, metraz_m2)) AS cena_m2,
        winda,
        COALESCE(
            SAFE_CAST(pietro_mieszkania AS INT64),
            SAFE_CAST(REGEXP_EXTRACT(pietro_normalized, r'^\s*(\d+)' AS INT64),
            SAFE_CAST(REGEXP_EXTRACT(pietro_raw,      r'^\s*(\d+)' AS INT64)
        ) AS pietro_num
    FROM `realstate-market-poland.prod.offers_sample_per_city`
    WHERE (cena_za_m2 IS NOT NULL OR (cena IS NOT NULL AND metraz_m2 IS NOT NULL AND metraz_m2 > 0))
),
filt AS (
    SELECT
        miasto,
        CASE WHEN pietro_num >= 10 THEN '10+' ELSE CAST(pietro_num AS STRING) END AS pietro_bucket,
        winda AS winda_yes,
        cena_m2
    FROM base
    WHERE pietro_num IS NOT NULL AND pietro_num >= 1
        AND winda IS NOT NULL
)

```

```

SELECT
miasto,
pietro_bucket,
COUNT(*) AS liczba_ofert,
COUNTIF(winda_yes) AS n_winda,
COUNTIF(NOT winda_yes) AS n_bez_windy,

-- średnie m2 (TAK/NIE)
ROUND(AVG(IF(winda_yes, cena_m2, NULL)), 0) AS avg_m2_winda,
ROUND(AVG(IF(NOT winda_yes, cena_m2, NULL)), 0) AS avg_m2_bez_windy,

-- premia (różnica TAK - NIE)
ROUND(AVG(IF(winda_yes, cena_m2, NULL)) - AVG(IF(NOT winda_yes, cena_m2, NULL)), 0) AS premia_avg_m2,
ROUND(100 * SAFE_DIVIDE(
    AVG(IF(winda_yes, cena_m2, NULL)) - AVG(IF(NOT winda_yes, cena_m2, NULL)),
    AVG(IF(NOT winda_yes, cena_m2, NULL))
), 1) AS premia_avg_pct,

-- mediany m2
ROUND(APPROX_QUANTILES(IF(winda_yes, cena_m2, NULL), 101)[OFFSET(50)], 0) AS med_m2_winda,
ROUND(APPROX_QUANTILES(IF(NOT winda_yes, cena_m2, NULL), 101)[OFFSET(50)], 0) AS med_m2_bez_windy,
ROUND(
    APPROX_QUANTILES(IF(winda_yes, cena_m2, NULL), 101)[OFFSET(50)]
    - APPROX_QUANTILES(IF(NOT winda_yes, cena_m2, NULL), 101)[OFFSET(50)]
, 0) AS premia_med_m2,

-- NOWE: średnia m2 niezależnie od windy
ROUND(AVG(cena_m2), 0) AS avg_m2_all
FROM filt
GROUP BY miasto, pietro_bucket
-- HAVING n_winda >= 10 AND n_bez_windy >= 10 -- progi liczebności
ORDER BY miasto, CASE WHEN pietro_bucket = '10+' THEN 10 ELSE CAST(pietro_bucket AS INT64) END;

```

Rynek pierwotny vs wtóry

```
WITH src AS (
    SELECT
        CASE
            WHEN LOWER(miasto) IN ('poznan','poznań') THEN 'Poznań'
            WHEN LOWER(miasto) IN ('kraków','krakow') THEN 'Kraków'
            WHEN LOWER(miasto) IN ('gdańsk','gdansk') THEN 'Gdańsk'
            WHEN LOWER(miasto) = 'warszawa' THEN 'Warszawa'
            WHEN LOWER(miasto) IN ('wrocław','wroclaw') THEN 'Wrocław'
            ELSE miasto
        END AS miasto,
        COALESCE(cena_za_m2, SAFE_DIVIDE(cena, metraz_m2)) AS cena_m2,
        rok_budowy,
        stan_wykonzenia,
        rodzaj_zabudowy
    FROM `realstate-market-poland.prod.offers_sample_per_city`
    WHERE (cena_za_m2 IS NOT NULL OR (cena IS NOT NULL AND metraz_m2 IS NOT NULL AND metraz_m2 > 0))
),
lab AS (
    SELECT
        miasto,
        CASE
            WHEN REGEXP_CONTAINS(LOWER(COALESCE(stan_wykonzenia,"")),
                r'deweloperski|do\s*wyko(n|ń)czenia|stan\s*surowy')
            OR (
                rok_budowy IS NOT NULL
                AND rok_budowy >= EXTRACT(YEAR FROM CURRENT_DATE()) - 2
                AND REGEXP_CONTAINS(LOWER(COALESCE(rodzaj_zabudowy,"")), r'apartamentowiec')
            )
            THEN 'pierwotny' ELSE 'wtóry'
        END AS rynek,
        cena_m2
    FROM src
    WHERE cena_m2 IS NOT NULL AND cena_m2 > 0
)
SELECT
    miasto,
```

```

rynek,
COUNT(*) AS liczba_ofert,
ROUND(AVG(cena_m2), 0) AS sr_cena_m2,
APPROX_QUANTILES(cena_m2, 101)[OFFSET(50)] AS mediana_cena_m2
FROM lab
GROUP BY miasto, rynek
ORDER BY miasto, rynek;

```

Sprawdzenie kompletności danych dla miast

```

WITH src AS (
    SELECT
        CASE
            WHEN LOWER(miasto) IN ('poznan','poznań') THEN 'Poznań'
            WHEN LOWER(miasto) IN ('kraków','krakow') THEN 'Kraków'
            WHEN LOWER(miasto) IN ('gdańsk','gdansk') THEN 'Gdańsk'
            WHEN LOWER(miasto) = 'warszawa' THEN 'Warszawa'
            WHEN LOWER(miasto) IN ('wrocław','wroclaw') THEN 'Wrocław'
            ELSE NULL
        END AS miasto_norm,
        s.*
    FROM `realstate-market-poland.prod.offers_sample_per_city` s
    WHERE LOWER(miasto) IN ('poznan','poznań','kraków','krakow','gdańsk','gdansk','warszawa','wrocław','wroclaw')
),
flat AS (
    SELECT
        miasto_norm AS miasto,
        i.kolumna,
        i.is_null
    FROM src s,
    UNNEST([
        STRUCT('offer_id' AS kolumna, s.offer_id IS NULL AS is_null),
        STRUCT('miasto' AS kolumna, s.miasto IS NULL OR TRIM(s.miasto) = "" AS is_null),
        STRUCT('typ_ogoszeniodawcy' AS kolumna, s.typ_ogoszeniodawcy IS NULL OR TRIM(s.typ_ogoszeniodawcy) = "" AS is_null),
        STRUCT('adres' AS kolumna, s.adres IS NULL OR TRIM(s.adres) = "" AS is_null)
    ])
)

```

```

STRUCT('adres_array'          AS kolumna, s.adres_array IS NULL OR ARRAY_LENGTH(s.adres_array) = 0
AS is_null),
STRUCT('cena'                 AS kolumna, s.cena IS NULL
AS is_null),
STRUCT('metraz_m2'            AS kolumna, s.metraz_m2 IS NULL
AS is_null),
STRUCT('czynsz'               AS kolumna, s.czynsz IS NULL
AS is_null),
STRUCT('stan_wykonczenia'     AS kolumna, s.stan_wykonczenia IS NULL OR TRIM(s.stan_wykonczenia) = ""
AS is_null),
STRUCT('ogrzewanie'           AS kolumna, s.ogrzewanie IS NULL OR TRIM(s.ogrzewanie) = ""
AS is_null),
STRUCT('taras_balkon'          AS kolumna, s.taras_balkon IS NULL
AS is_null),
STRUCT('winda'                AS kolumna, s.winda IS NULL
AS is_null),
STRUCT('rok_budowy'            AS kolumna, s.rok_budowy IS NULL
AS is_null),
STRUCT('rodzaj_zabudowy'       AS kolumna, s.rodzaj_zabudowy IS NULL OR TRIM(s.rodzaj_zabudowy) = ""
AS is_null),
STRUCT('material_budynku'      AS kolumna, s.material_budynku IS NULL OR TRIM(s.material_budynku) = ""
AS is_null),
STRUCT('media'                 AS kolumna, s.media IS NULL OR ARRAY_LENGTH(s.media) = 0
AS is_null),
STRUCT('zabezpieczenia'        AS kolumna, s.zabezpieczenia IS NULL OR ARRAY_LENGTH(s.zabezpieczenia) = 0
AS is_null),
STRUCT('wyposazenie'           AS kolumna, s.wyposazenie IS NULL OR ARRAY_LENGTH(s.wyposazenie) = 0
AS is_null),
STRUCT('certyfikat_energetyczny' AS kolumna, s.certyfikat_energetyczny IS NULL OR
TRIM(s.certyfikat_energetyczny) = "" AS is_null),
STRUCT('informacje_dodatkowe'    AS kolumna, s.informacje_dodatkowe IS NULL OR
ARRAY_LENGTH(s.informacje_dodatkowe)=0 AS is_null),
STRUCT('cena_za_m2'              AS kolumna, s.cena_za_m2 IS NULL
AS is_null),
STRUCT('pietro_mieszkania'       AS kolumna, s.pietyro_mieszkania IS NULL
AS is_null),
STRUCT('najwyzsze_pietro'         AS kolumna, s.najwyzsze_pietro IS NULL
AS is_null),
STRUCT('pietro_raw'              AS kolumna, s.pietyro_raw IS NULL OR TRIM(s.pietyro_raw) = ""
AS is_null),
STRUCT('pietro_normalized'        AS kolumna, s.pietyro_normalized IS NULL OR TRIM(s.pietyro_normalized) = ""
AS is_null),
STRUCT('procent_wysokosci_budynku' AS kolumna, s.procent_wysokosci_budynku IS NULL
AS is_null),
STRUCT('garazmiejsc_parkingowe'   AS kolumna, s.garazmiejsc_parkingowe IS NULL
AS is_null),
STRUCT('piwnica'                 AS kolumna, s.piwnica IS NULL
AS is_null)
]) AS i
)
SELECT
miasto,
kolumna,
COUNT(*)          AS liczba_wierszy,

```

```

SUM(CASE WHEN is_null THEN 1 ELSE 0 END)           AS liczba_null,
COUNT(*) - SUM(CASE WHEN is_null THEN 1 ELSE 0 END)   AS liczba_poprawnych,
ROUND(100 * SAFE_DIVIDE(SUM(CASE WHEN is_null THEN 1 END), COUNT(*)), 2) AS null_pct
FROM flat
GROUP BY miasto, kolumna
ORDER BY miasto, null_pct DESC, kolumna;

```

Zależność ceny od typu ogrzewania

```

WITH src AS (
SELECT
CASE
WHEN LOWER(miasto) IN ('poznan','poznań') THEN 'Poznań'
WHEN LOWER(miasto) IN ('kraków','krakow') THEN 'Kraków'
WHEN LOWER(miasto) IN ('gdańsk','gdansk') THEN 'Gdańsk'
WHEN LOWER(miasto) = 'warszawa'      THEN 'Warszawa'
WHEN LOWER(miasto) IN ('wrocław','wroclaw') THEN 'Wrocław'
ELSE miasto
END AS miasto,
LOWER(TRIM(ogrzewanie)) AS ogrzewanie,
COALESCE(cena_za_m2, SAFE_DIVIDE(cena, metraz_m2)) AS cena_m2
FROM `realstate-market-poland.prod.offers_sample_per_city`
WHERE (cena_za_m2 IS NOT NULL OR (cena IS NOT NULL AND metraz_m2 > 0))
)
SELECT
miasto, ogrzewanie,
COUNT(*) AS n,
ROUND(AVG(cena_m2),0) AS sr_m2,
ROUND(APPROX_QUANTILES(cena_m2,101)[OFFSET(50)],0) AS mediana_m2
FROM src
WHERE ogrzewanie IS NOT NULL AND ogrzewanie <> ""
GROUP BY miasto, ogrzewanie
HAVING n >= 10
ORDER BY miasto, sr_m2 DESC;

```

Udział procentowy mieszkań w bazie danych zbudowane w przedziałach lat

```
WITH src AS (
    SELECT
        CASE
            WHEN LOWER(miasto) IN ('poznan','poznań') THEN 'Poznań'
            WHEN LOWER(miasto) IN ('kraków','krakow') THEN 'Kraków'
            WHEN LOWER(miasto) IN ('gdańsk','gdansk') THEN 'Gdańsk'
            WHEN LOWER(miasto) = 'warszawa' THEN 'Warszawa'
            WHEN LOWER(miasto) IN ('wrocław','wroclaw') THEN 'Wrocław'
            ELSE miasto
        END AS miasto,
        rok_budowy,
        metraz_m2,
        COALESCE(cena_za_m2, SAFE_DIVIDE(cena, metraz_m2)) AS cena_m2
    FROM `realstate-market-poland.prod.offers_sample_per_city`
)
SELECT
    miasto,
    COUNT(*) AS liczba_ofert,
    -- Udziały % liczby ofert
    ROUND(100 * SAFE_DIVIDE(SUM(CASE WHEN rok_budowy < 1940 THEN 1 ELSE 0 END), COUNT(*)), 1) AS pct_lt_1940,
    ROUND(100 * SAFE_DIVIDE(SUM(CASE WHEN rok_budowy BETWEEN 1940 AND 1960 THEN 1 ELSE 0 END), COUNT(*)), 1) AS pct_1940_1960,
    ROUND(100 * SAFE_DIVIDE(SUM(CASE WHEN rok_budowy BETWEEN 1960 AND 1980 THEN 1 ELSE 0 END), COUNT(*)), 1) AS pct_1960_1980,
    ROUND(100 * SAFE_DIVIDE(SUM(CASE WHEN rok_budowy BETWEEN 1980 AND 2000 THEN 1 ELSE 0 END), COUNT(*)), 1) AS pct_1980_2000,
    ROUND(100 * SAFE_DIVIDE(SUM(CASE WHEN rok_budowy BETWEEN 2000 AND 2020 THEN 1 ELSE 0 END), COUNT(*)), 1) AS pct_2000_2020,
    ROUND(100 * SAFE_DIVIDE(SUM(CASE WHEN rok_budowy > 2020 THEN 1 ELSE 0 END), COUNT(*)), 1) AS pct_gt_2020,
    ROUND(100 * SAFE_DIVIDE(SUM(CASE WHEN rok_budowy IS NULL THEN 1 ELSE 0 END), COUNT(*)), 1) AS pct_brak_danych,
    -- Średnia cena za m2 w koszykach
)
```

-- Średnia cena za m² w koszykach

```

ROUND(AVG(CASE WHEN rok_budowy < 1940 THEN cena_m2 END), 0) AS avg_m2_lt_1940,
ROUND(AVG(CASE WHEN rok_budowy BETWEEN 1940 AND 1960 THEN cena_m2 END), 0) AS
avg_m2_1940_1960,
ROUND(AVG(CASE WHEN rok_budowy BETWEEN 1960 AND 1980 THEN cena_m2 END), 0) AS
avg_m2_1960_1980,
ROUND(AVG(CASE WHEN rok_budowy BETWEEN 1980 AND 2000 THEN cena_m2 END), 0) AS
avg_m2_1980_2000,
ROUND(AVG(CASE WHEN rok_budowy BETWEEN 2000 AND 2020 THEN cena_m2 END), 0) AS
avg_m2_2000_2020,
ROUND(AVG(CASE WHEN rok_budowy > 2020 THEN cena_m2 END), 0) AS avg_m2_gt_2020,
ROUND(AVG(CASE WHEN rok_budowy IS NULL THEN cena_m2 END), 0) AS
avg_m2_brak_danych,

```

-- średni metraż (m²) w koszykach

```

ROUND(AVG(CASE WHEN rok_budowy < 1940 THEN metraz_m2 END), 1) AS avg_metraz_lt_1940,
ROUND(AVG(CASE WHEN rok_budowy BETWEEN 1940 AND 1960 THEN metraz_m2 END), 1) AS
avg_metraz_1940_1960,
ROUND(AVG(CASE WHEN rok_budowy BETWEEN 1960 AND 1980 THEN metraz_m2 END), 1) AS
avg_metraz_1960_1980,
ROUND(AVG(CASE WHEN rok_budowy BETWEEN 1980 AND 2000 THEN metraz_m2 END), 1) AS
avg_metraz_1980_2000,
ROUND(AVG(CASE WHEN rok_budowy BETWEEN 2000 AND 2020 THEN metraz_m2 END), 1) AS
avg_metraz_2000_2020,
ROUND(AVG(CASE WHEN rok_budowy > 2020 THEN metraz_m2 END), 1) AS avg_metraz_gt_2020,
ROUND(AVG(CASE WHEN rok_budowy IS NULL THEN metraz_m2 END), 1) AS
avg_metraz_brak_danych

```

FROM src
GROUP BY miasto
ORDER BY miasto;

Udział procentowy mieszkań w przedziałach metrażowych

```

WITH src AS (
SELECT
CASE
WHEN LOWER(miasto) IN ('poznan','poznań') THEN 'Poznań'
WHEN LOWER(miasto) IN ('kraków','krakow') THEN 'Kraków'
WHEN LOWER(miasto) IN ('gdańsk','gdansk') THEN 'Gdańsk'
WHEN LOWER(miasto) = 'warszawa' THEN 'Warszawa'

```

```

WHEN LOWER(miasto) IN ('wrocław','wroclaw') THEN 'Wrocław'
ELSE NULL
END AS miasto,
metraz_m2,
COALESCE(cena_za_m2, SAFE_DIVIDE(cena, metraz_m2)) AS cena_m2
FROM `realstate-market-poland.prod.offers_sample_per_city`
WHERE LOWER(miasto) IN ('poznan','poznań','kraków','krakow','gdańsk','gdansk','warszawa','wrocław','wroclaw')
AND metraz_m2 IS NOT NULL AND metraz_m2 > 0
AND (cena_za_m2 IS NOT NULL OR (cena IS NOT NULL AND cena > 0))
)
SELECT
miasto,
COUNT(*) AS liczba_ofert,
-- % udział liczby ofert
ROUND(100 * AVG(CASE WHEN metraz_m2 < 25           THEN 1 ELSE 0 END), 1) AS pct_lt_25,
ROUND(100 * AVG(CASE WHEN metraz_m2 >= 25 AND metraz_m2 < 40 THEN 1 ELSE 0 END), 1) AS pct_25_40,
ROUND(100 * AVG(CASE WHEN metraz_m2 >= 40 AND metraz_m2 < 60 THEN 1 ELSE 0 END), 1) AS pct_40_60,
ROUND(100 * AVG(CASE WHEN metraz_m2 >= 60 AND metraz_m2 < 80 THEN 1 ELSE 0 END), 1) AS pct_60_80,
ROUND(100 * AVG(CASE WHEN metraz_m2 >= 80           THEN 1 ELSE 0 END), 1) AS pct_ge_80,
-- średnia cena za m2 w przedziałach
ROUND(AVG(CASE WHEN metraz_m2 < 25           THEN cena_m2 END), 0) AS sr_m2_lt_25,
ROUND(AVG(CASE WHEN metraz_m2 >= 25 AND metraz_m2 < 40 THEN cena_m2 END), 0) AS sr_m2_25_40,
ROUND(AVG(CASE WHEN metraz_m2 >= 40 AND metraz_m2 < 60 THEN cena_m2 END), 0) AS sr_m2_40_60,
ROUND(AVG(CASE WHEN metraz_m2 >= 60 AND metraz_m2 < 80 THEN cena_m2 END), 0) AS sr_m2_60_80,
ROUND(AVG(CASE WHEN metraz_m2 >= 80           THEN cena_m2 END), 0) AS sr_m2_ge_80
FROM src
GROUP BY miasto
ORDER BY miasto;

```

Udział procentowy mieszkań w stanach wykończenia

WITH src AS (

```

SELECT
CASE
    WHEN LOWER(miasto) IN ('poznan','poznań') THEN 'Poznań'
    WHEN LOWER(miasto) IN ('kraków','krakow') THEN 'Kraków'
    WHEN LOWER(miasto) IN ('gdańsk','gdansk') THEN 'Gdańsk'
    WHEN LOWER(miasto) = 'warszawa' THEN 'Warszawa'
    WHEN LOWER(miasto) IN ('wrocław','wroclaw') THEN 'Wrocław'
    ELSE miasto
END AS miasto,
CASE
    WHEN REGEXP_CONTAINS(LOWER(COALESCE(stan_wykonczenia,"")),
        r'pod\s*klucz|poł\s*remont|do\s*zamieszkania|wyko(n|ń)czo|gotow') THEN 'do_zamieszkania'
    WHEN REGEXP_CONTAINS(LOWER(COALESCE(stan_wykonczenia,"")),
        r'do\s*remontu|do\s*odświe|od\s*świe') THEN 'do_remontu'
    WHEN REGEXP_CONTAINS(LOWER(COALESCE(stan_wykonczenia,"")),
        r'dewelopersk|do\s*wyko(n|ń)czenia|stan\s*surowy') THEN 'do_wykonczenia'
    ELSE 'inne'
END AS stan_segment,
COALESCE(cena_za_m2, SAFE_DIVIDE(cena, metraz_m2)) AS cena_m2
FROM `realstate-market-poland.prod.offers_sample_per_city`
WHERE LOWER(miasto) IN ('poznan','poznań','kraków','krakow','gdańsk','gdansk','warszawa','wrocław','wroclaw')
    AND (cena_za_m2 IS NOT NULL OR (cena IS NOT NULL AND metraz_m2 IS NOT NULL AND metraz_m2 > 0))
),
agg AS (
SELECT
    miasto,
    stan_segment,
    COUNT(*) AS liczba_ofert,
    ROUND(AVG(cena_m2), 0) AS sr_cena_m2
FROM src
GROUP BY miasto, stan_segment
)
SELECT
    miasto,
    stan_segment,
    liczba_ofert,
    ROUND(
        100 * SAFE_DIVIDE(

```

```

liczba_ofert,
SUM(liczba_ofert) OVER (PARTITION BY miasto)
),
) AS udzial_pct,
sr_cena_m2
FROM agg
ORDER BY miasto, stan_segment;

```

Udział rynkowy danych wysokości, średnia cena za metr, pct winda

```

WITH src AS (
SELECT
CASE
WHEN LOWER(miasto) IN ('poznan','poznań') THEN 'Poznań'
WHEN LOWER(miasto) IN ('kraków','krakow') THEN 'Kraków'
WHEN LOWER(miasto) IN ('gdańsk','gdansk') THEN 'Gdańsk'
WHEN LOWER(miasto) = 'warszawa' THEN 'Warszawa'
WHEN LOWER(miasto) IN ('wrocław','wroclaw') THEN 'Wrocław'
ELSE miasto
END AS miasto,
COALESCE(
najwyzsze_pietro,
SAFE_CAST(REGEXP_EXTRACT(pietro_normalized, r'\s*(\d+)' AS INT64),
SAFE_CAST(REGEXP_EXTRACT(pietro_raw, r'\s*(\d+)' AS INT64)
) AS h,
winda,
COALESCE(cena_za_m2, SAFE_DIVIDE(cena, metraz_m2)) AS cena_m2
FROM `realstate-market-poland.prod.offers_sample_per_city`
),
bin AS (
SELECT
miasto,
h,
CASE
WHEN h BETWEEN 1 AND 3 THEN '1-3'
WHEN h BETWEEN 4 AND 5 THEN '4-5'

```

```

WHEN h BETWEEN 6 AND 8 THEN '6-8'
WHEN h BETWEEN 9 AND 12 THEN '9-12'
WHEN h >= 13      THEN '13+'
END AS wys_budynek,
winda,
cena_m2
FROM src
WHERE h IS NOT NULL AND h >= 1
AND cena_m2 IS NOT NULL
),
tot AS (
SELECT miasto, COUNT(*) AS n FROM bin GROUP BY miasto
),
city_max AS (
SELECT miasto, MAX(h) AS najwyzszy_budynek FROM bin GROUP BY miasto
)
SELECT
b.miasto,
cm.najwyzszy_budynek,
b.wys_budynek,
COUNT(*) AS liczba_ofert,
ROUND(100 * SAFE_DIVIDE(COUNT(*), t.n), 2) AS pct_udzial,
COUNTIF(b.winda IS NOT NULL)           AS n_winda_known,
COUNTIF(b.winda IS TRUE)               AS n_winda_true,
ROUND(100 * SAFE_DIVIDE(COUNTIF(b.winda IS TRUE),
                         COUNTIF(b.winda IS NOT NULL)), 2) AS pct_winda,
ROUND(AVG(IF(b.winda IS TRUE, b.cena_m2, NULL)), 0) AS sr_cena_winda,
ROUND(AVG(IF(b.winda IS FALSE, b.cena_m2, NULL)), 0) AS sr_cena_bez_windy
FROM bin b
JOIN tot t   USING (miasto)
JOIN city_max cm USING (miasto)
GROUP BY b.miasto, cm.najwyzszy_budynek, b.wys_budynek, t.n
ORDER BY b.miasto,
CASE b.wys_budynek
  WHEN '1-3' THEN 1 WHEN '4-5' THEN 2 WHEN '6-8' THEN 3
  WHEN '9-12' THEN 4 ELSE 5 END;

```

Średnia cena za metr a materiał budynku

```
WITH src AS (
    SELECT
        CASE
            WHEN LOWER(miasto) IN ('poznan','poznań') THEN 'Poznań'
            WHEN LOWER(miasto) IN ('kraków','krakow') THEN 'Kraków'
            WHEN LOWER(miasto) IN ('gdańsk','gdansk') THEN 'Gdańsk'
            WHEN LOWER(miasto) = 'warszawa' THEN 'Warszawa'
            WHEN LOWER(miasto) IN ('wrocław','wroclaw') THEN 'Wrocław'
            ELSE miasto
        END AS miasto,
        LOWER(TRIM(material_budynku)) AS material,
        COALESCE(cena_za_m2, SAFE_DIVIDE(cena, metraz_m2)) AS cena_m2
    FROM `realstate-market-poland.prod.offers_sample_per_city`
    WHERE (cena_za_m2 IS NOT NULL OR (cena IS NOT NULL AND metraz_m2 > 0))
)
SELECT
    miasto, material,
    COUNT(*) AS n,
    ROUND(AVG(cena_m2),0) AS sr_m2
FROM src
WHERE material IS NOT NULL
GROUP BY miasto, material
HAVING n >= 10
ORDER BY miasto, material;
```

Średnia cena za metr a typ budynku

```
WITH src AS (
    SELECT
        CASE
            WHEN LOWER(miasto) IN ('poznan','poznań') THEN 'Poznań'
            WHEN LOWER(miasto) IN ('kraków','krakow') THEN 'Kraków'
            WHEN LOWER(miasto) IN ('gdańsk','gdansk') THEN 'Gdańsk'
            WHEN LOWER(miasto) = 'warszawa' THEN 'Warszawa'
            WHEN LOWER(miasto) IN ('wrocław','wroclaw') THEN 'Wrocław'
```

```
ELSE miasto
END AS miasto,
LOWER(TRIM(rodzaj_zabudowy)) AS rodzaj,
COALESCE(cena_za_m2, SAFE_DIVIDE(cena, metraz_m2)) AS cena_m2
FROM `realstate-market-poland.prod.offers_sample_per_city`
WHERE (cena_za_m2 IS NOT NULL OR (cena IS NOT NULL AND metraz_m2 > 0))
)
SELECT
miasto, rodzaj,
COUNT(*) AS n,
ROUND(AVG(cena_m2),0) AS sr_m2
FROM src
WHERE rodzaj IS NOT NULL
GROUP BY miasto, rodzaj
HAVING n >= 10
ORDER BY miasto, rodzaj;
```