

Prezentacja EDA danych historycznych ofert pracy z portalu JustJoinIT

Rozglądając się za interesującymi mnie ofertami zrobiłem analize archiwalnych ofert z ostatnich lat

Zapraszam do zapoznania się z moją analizą.

1. Ogólny przegląd danych.

Dane zawierają 8725555 wierszy i 24 kolumn.

Wyświetlenie 10 losowych rekordów

	title	street	city	country_code	address_text	marker_icon	workplace_type	company_name	
8473172	Specjalista SAP S/4HANA	Centrum	Katowice	PL	Centrum, Katowice	erp	remote	INWEMER	ht
339854	C#/.NET Software Engineer	Aleja Grunwaldzka 415	Gdańsk	PL	Aleja Grunwaldzka 415, Gdańsk	c	partly_remote	Lufthansa Systems	h
1346821	QA Engineer	Centrum	Warszawa	PL	Centrum, Warszawa	testing	partly_remote	Evolution	https://www.evolution.com/car
4474412	Senior Game Tester	-	Warszawa	PL	-, Warszawa	testing	remote	Vivid Games	https/,
342508	Java Developer	Centrum	Poznań	PL	Centrum, Poznań	java	remote	Idego	
7068435	Cloud Security Engineer	-	Bydgoszcz	PL	-, Bydgoszcz	security	remote	PwC Polska	htt
684724	Senior Frontend Developer	Ligocka 103/budynek 8	Katowice	PL	Ligocka 103/budynek 8, Katowice	javascript	remote	Proexe Sp. z o.o.	I
8721232	Senior Fullstack Developer	-	Łódź	PL	-, Łódź	net	remote	Cavendish Professionals	http://www.cavendishj
6459735	C# Developer	Centrum	Bytom	PL	Centrum, Bytom	net	partly_remote	SoftTech	
2405058	Senior Java Developer w zespole xREF	Centrum	Bydgoszcz	PL	Centrum, Bydgoszcz	java	remote	XTB	https,,

Sprawdzenie jakie mamy kolumny & typy danych

```
Index(['title', 'street', 'city', 'country_code', 'address_text',
      'marker_icon', 'workplace_type', 'company_name', 'company_url',
      'company_size', 'experience_level', 'latitude', 'longitude',
      'published_at', 'remote_interview', 'id', 'employment_types',
      'company_logo_url', 'skills', 'remote', 'open_to_hire_ukrainians',
      'display_offer', 'multilocation', 'way_of_apply'],
      dtype='object')
```

10 entries per page

Search:

	0
title	object
street	object
city	object
country_code	object
address_text	object
marker_icon	object
workplace_type	object
company_name	object
company_url	object
company_size	object

Showing 1 to 10 of 24 entries

Sprawdzenie podsumowania dla kolumn numerycznych oraz wartości unikalne

	latitude	longitude
count	8725555	8725555
mean	51.65359	18.59925
std	2.610446	8.528313
min	-42.9758	-123.1207
25%	50.66832	17.03854
50%	52.12052	19.40607
75%	52.40936	20.98584
max	65.01209	174.886

	0 ▴ ▾
title	33818
street	10316
city	1650
country_code	88
address_text	13734
marker_icon	24
workplace_type	3
company_name	7676
company_url	8180
company_size	1108

Showing 1 to 10 of 24 entries

Obserwacje wstępne:

1. Dane archiwalne ofert pracy ze strony JustJoinIT.
2. Dane zawierają **8725555** wierszy oraz **24** kolumn.

2. Przygotowanie danych

Usuwanie kolumn & naprawa kolumny z datami & usuwanie duplikatów

	0 ▴ ▾
title	object
street	object
city	object
country_code	object
marker_icon	object
workplace_type	object
company_name	object
experience_level	object
published_at	datetime64[ns, UTC]
remote_interview	bool

Showing 1 to 10 of 15 entries

	0 ▴ ▾
title	object
street	object
city	object
country_code	object
marker_icon	object
workplace_type	object
company_name	object
experience_level	object
published_at	datetime64[ns, UTC]
remote_interview	bool

Showing 1 to 10 of 15 entries

	title	street	city	country_code	marker_icon	workplace_type	company_name	ex
14	Full Stack .Net Developer	Baletowa 1	Warszawa	PL	net	remote	Novaris Sp. z o.o.	
16	Product Designer	Masnedøgade 22	København	DK	ux	remote	Contractbook	
28	Python Developer with Django	Centrum	Katowice	PL	python	remote	Codahead	
39	MLOps Engineer	Mikołaja Kopernika 95a	Białystok	PL	python	partly_remote	Grape Up	
47	Fullstack Developer (Java)	Strzegomska 142A	Wrocław	PL	java	remote	Objectivity	
49	MLOps Engineer	Fabryczna 6	Wrocław	PL	python	partly_remote	Grape Up	
51	Regular/Senior JAVA Developer Chatbot	Centrum	Kraków	PL	java	remote	JCommerce	
59	Mid Java Developer	Centrum	Wrocław	PL	java	remote	Novaris Sp. z o.o.	
60	Analityk Systemowy (Integracje Systemów)	Centrum	Kraków	PL	analytics	remote	Unity Group	
64	Analityk Systemowy (Integracje Systemów)	Centrum	Warszawa	PL	analytics	remote	Unity Group	

	title	street	city	country_code	marker_icon	workplace_type	company_name	experience_level	published_at	remote
5	Analityk Systemowy (Integracje Systemów)	Strzegomska 2-4	Wrocław	PL	analytics	remote	Unity Group	senior	2021-10-23 08:00:23.064000+00:00	
60	Analityk Systemowy (Integracje Systemów)	Centrum	Kraków	PL	analytics	remote	Unity Group	senior	2021-10-23 08:00:23.064000+00:00	
64	Analityk Systemowy (Integracje Systemów)	Centrum	Warszawa	PL	analytics	remote	Unity Group	senior	2021-10-23 08:00:23.064000+00:00	
7545	Analityk Systemowy (Integracje Systemów)	Centrum	Kraków	PL	analytics	remote	Unity Group	senior	2021-10-23 08:00:23.064000+00:00	
7546	Analityk Systemowy (Integracje Systemów)	Centrum	Warszawa	PL	analytics	remote	Unity Group	senior	2021-10-23 08:00:23.064000+00:00	
7553	Analityk Systemowy (Integracje Systemów)	Strzegomska 2-4	Wrocław	PL	analytics	remote	Unity Group	senior	2021-10-23 08:00:23.064000+00:00	
14977	Analityk Systemowy (Integracje Systemów)	Strzegomska 2-4	Wrocław	PL	analytics	remote	Unity Group	senior	2021-10-23 08:00:23.064000+00:00	
15029	Analityk Systemowy (Integracje Systemów)	Centrum	Kraków	PL	analytics	remote	Unity Group	senior	2021-10-23 08:00:23.064000+00:00	
15035	Analityk Systemowy (Integracje Systemów)	Centrum	Warszawa	PL	analytics	remote	Unity Group	senior	2021-10-23 08:00:23.064000+00:00	
22521	Analityk Systemowy (Integracje Systemów)	Strzegomska 2-4	Wrocław	PL	analytics	remote	Unity Group	senior	2021-10-23 08:00:23.064000+00:00	

(2096717, 15)

Przeprowadzone operacje wstępne:

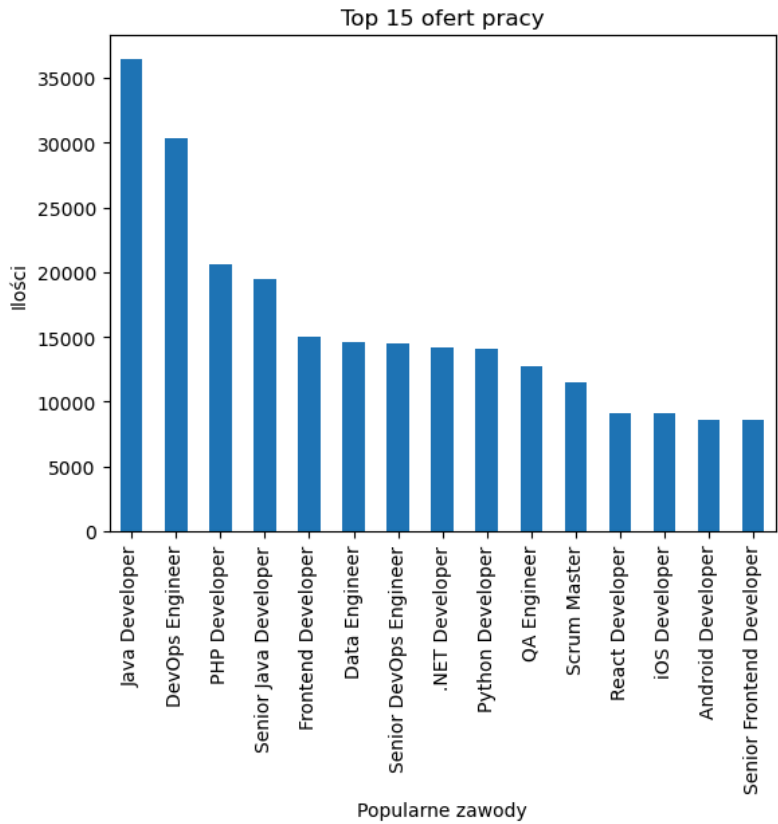
1. Sprawdzenie jakie dane zawiera DataFrame oraz usunięcie zbędnych kolumn.
2. Naprawienie kolumny z czasem publikacji.
3. Wykrycie dużej ilości duplikatów, usunięcie powtarzających się ofert pracy w danych.

3. Przegląd pojedynczysz kolumn

1. Stanowiska pracy i TOP15 poszukiwanych pracowników.

	count
title	
Java Developer	36445
DevOps Engineer	30289
PHP Developer	20595
Senior Java Developer	19491
Frontend Developer	15021
Data Engineer	14562
Senior DevOps Engineer	14533
.NET Developer	14184
Python Developer	14129
QA Engineer	12767

Text(0, 0.5, 'Ilości')



Obserwacje:

1. Dane zawierają: **33818** stanowisk pracy z: **7676** firm.

2. Pierwsza piętnastka najczęściej występujących ofert:

- Java Developer: 36445
- DevOps Engineer: 30289
- PHP Developer: 20595
- Senior Java Developer: 19491
- Frontend Developer: 15021
- Data Engineer: 14562
- Senior DevOps Engineer: 14533
- .NET Developer: 14184
- Python Developer: 14129
- QA Engineer: 12767
- Scrum Master: 11485
- React Developer: 9091
- iOS Developer: 9059
- Android Developer: 8571
- Senior Frontend Developer: 8543.

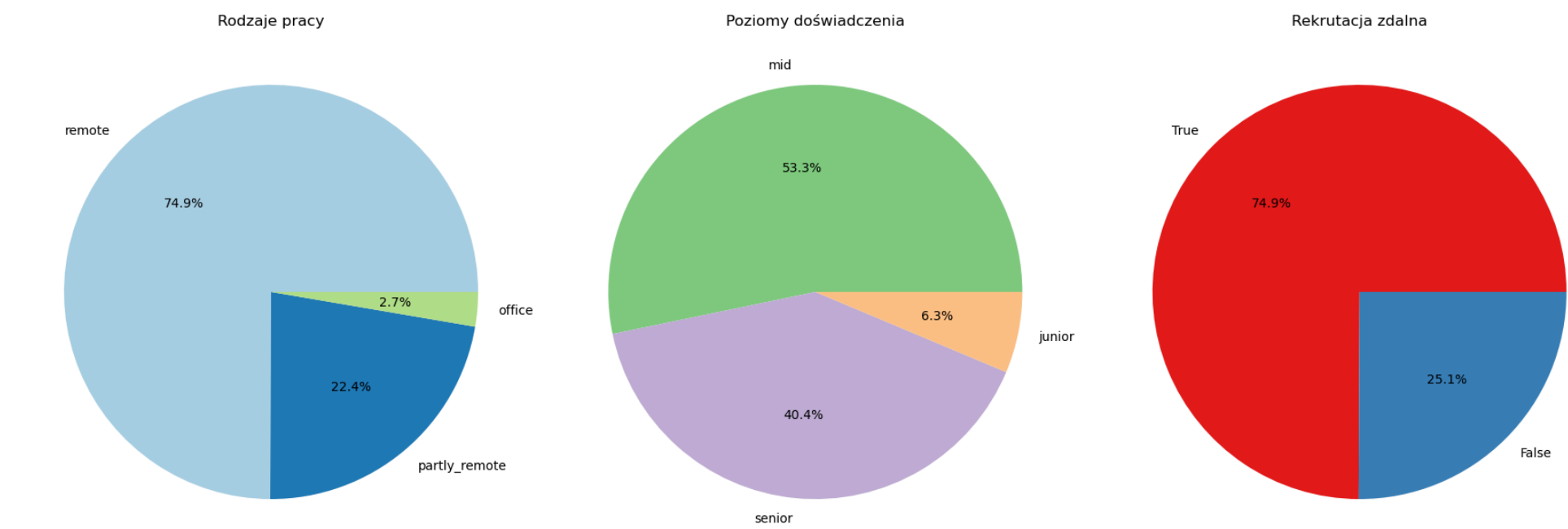
2. Typy pracy, poziom doświadczenia oraz rekrutacja zdalna.

	count
workplace_type	
remote	1571153
partly_remote	469233
office	56331

	count
experience_level	
mid	1116550
senior	847425
junior	132742

	count
remote	
true	1571153
false	525564

Wykresy



Wnioski:

1. Dane zawierają: 3 rodzaje pracy:

- remote: 1571153 (74.93%)
- partly_remote: 469233 (22.38%)
- office: 56331 (2.69%).

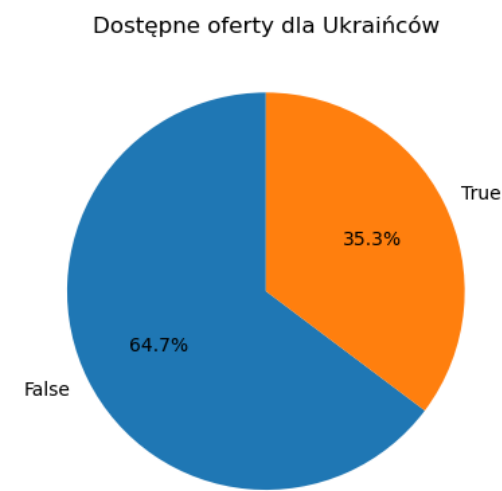
2. Liczba ofert pracy dla poszczególnych poziomów doświadczenia:

- mid: 1116550 (53.25%)
- senior: 847425 (40.42%)
- junior: 132742 (6.33%).

3. Liczba ofert pracy z możliwością rekrutacji zdalnej:

- True: 1571153 (74.93%)
- False: 525564 (25.07%).

3. Sprawdzenie częstotliwości ofert pracy dla Ukraińców.



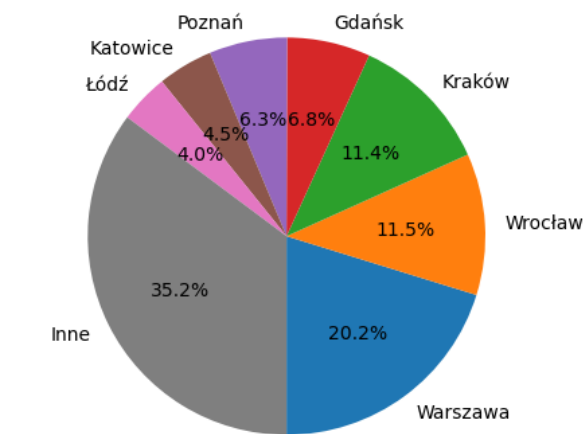
Obserwacje:

Oferty dla Ukraińców:

- False: 1227127 (64.74%)
- True: 668239 (35.26%).

4. Analiza ofert pracy w poszczególnych miastach.

Oferty zatrudnienia w poszczególnych miastach



Obserwacje:

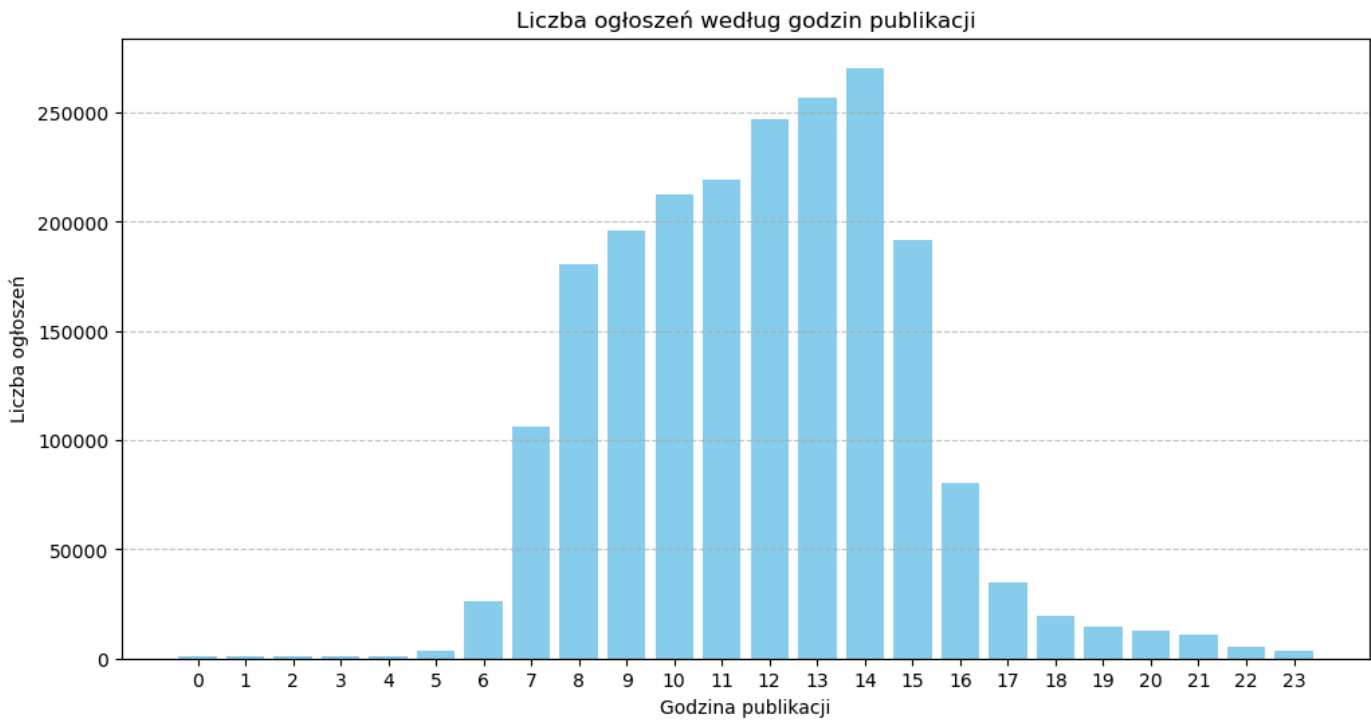
Prezentacja ilości ofert pracy w danych miejscowościach i ich procent w danych:

- Warszawa: 423917 (20.22%)
- Wrocław: 241610 (11.52%)
- Kraków: 239869 (11.44%)
- Gdańsk: 142087 (6.78%)
- Poznań: 132900 (6.34%)
- Katowice: 94199 (4.49%)
- Łódź: 83936 (4.00%)
- Szczecin: 52732 (2.51%)
- Lublin: 50519 (2.41%)
- Bydgoszcz: 47772 (2.28%)
- Rzeszów: 46212 (2.20%)
- Białystok: 45668 (2.18%).

W danych historycznych możemy zaobserwować że największe zapotrzebowanie na specjalistów w branży IT jest w miejscowościach:

- Warszawa
- Wrocław
- Kraków

5. Analiza czasu publikacji ogłoszeń.

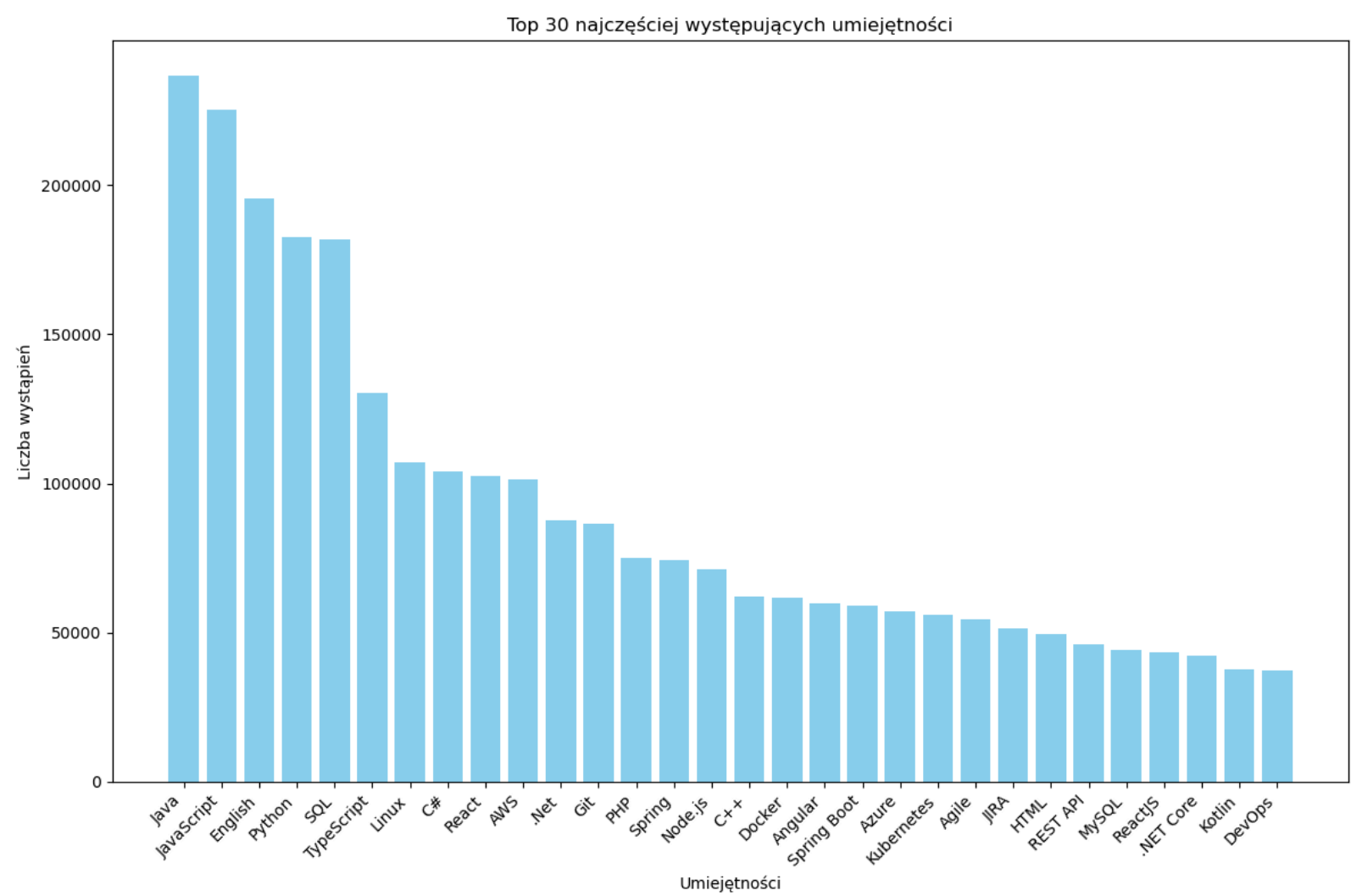


Obserwacje:

- Dane historyczne zamieszczane były w godzinach pracy między 7 a 16.
- Ta obserwacja prowadzi nas do pytania, czy oferty są sprawdzane przez administratorów/moderatorów i dodawane w ich godzinach pracy?
- W godzinach wieczornych między 17-6 rano widzimy minimalną aktywność publikacji ofert, co może nam sugerować dodawanie ofert np. zagranicznych.

6. Analiza wymaganych umiejętności oraz sposób zatrudnienia.

1. Umiejętności



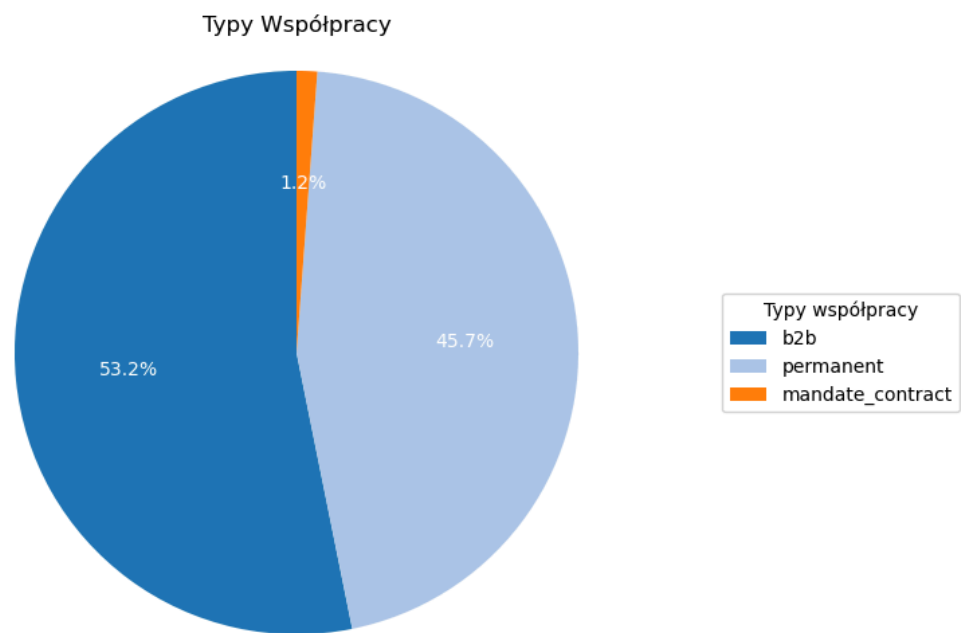
Obserwacje:

1. Ilość unikalnych umiejętności w danych: 11193

2. Trzydzieści najczęściej wymienianych umiejętności w ofertach pracy:

- Java: 236559 (3.96%)
- JavaScript: 225252 (3.77%)
- English: 195453 (3.27%)
- Python: 182611 (3.05%)
- SQL: 181772 (3.04%)
- TypeScript: 130448 (2.18%)
- Linux: 107132 (1.79%)
- C#: 103838 (1.74%)
- React: 102671 (1.72%)
- AWS: 101340 (1.70%)
- .Net: 87792 (1.47%)
- Git: 86587 (1.45%)
- PHP: 75092 (1.26%)
- Spring: 74265 (1.24%)
- Node.js: 71379 (1.19%)
- C++: 62088 (1.04%)
- Docker: 61755 (1.03%)
- Angular: 59639 (1.00%)
- Spring Boot: 59170 (0.99%)
- Azure: 57224 (0.96%)
- Kubernetes: 56084 (0.94%)
- Agile: 54408 (0.91%)
- JIRA: 51383 (0.86%)
- HTML: 49555 (0.83%)
- REST API: 46187 (0.77%)
- MySQL: 44099 (0.74%)
- ReactJS: 43377 (0.73%)
- .NET Core: 42088 (0.70%)
- Kotlin: 37539 (0.63%)
- DevOps: 37268 (0.62%)

2. Sposób zatrudnienia / współpracy

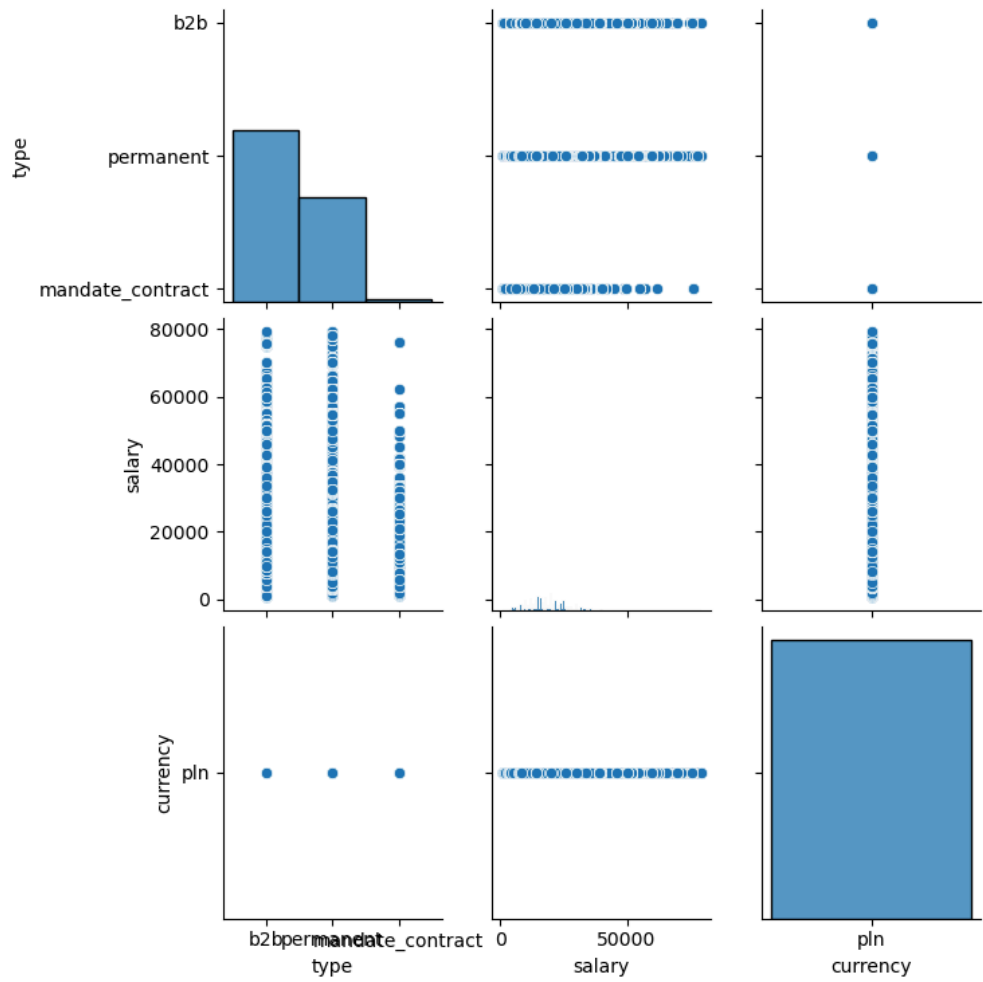
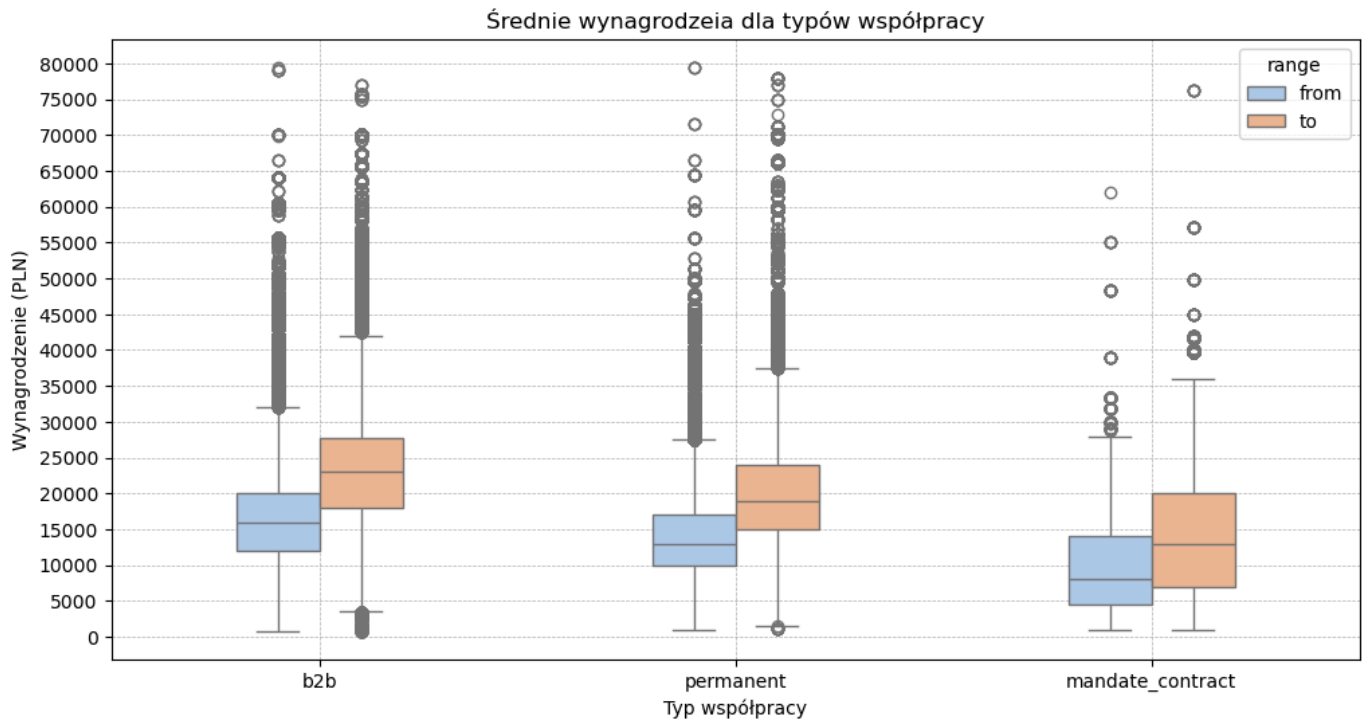
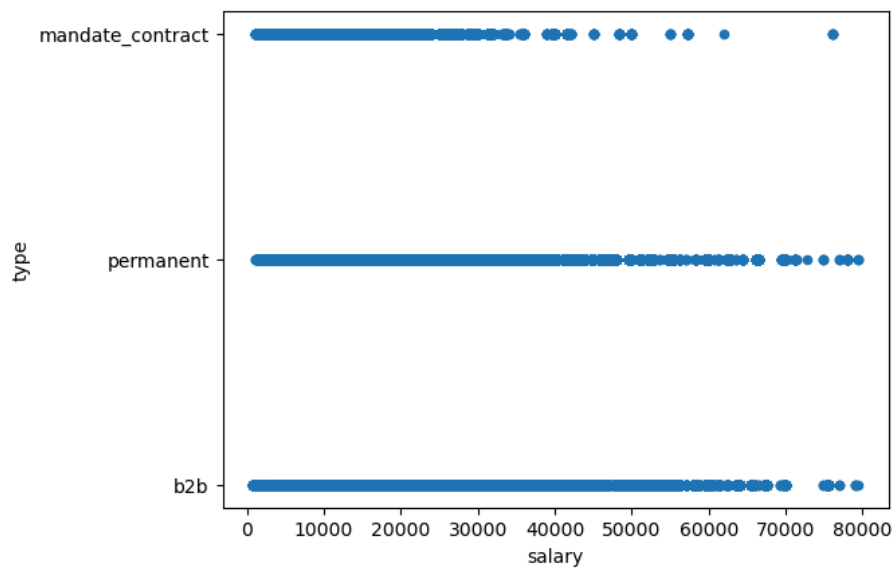


Obserwacje dotyczące sposobów współpracy:

Najczęściej wymieniane sposoby współpracy:

- b2b: 1431176 (53.2%)
- permanent: 1228847 (45.7%)
- mandate_contract: 31720 (1.2%).

7. Analiza wynagrodzeń



Obserwacje wynagrodzeń we wszystkich typach współpracy:

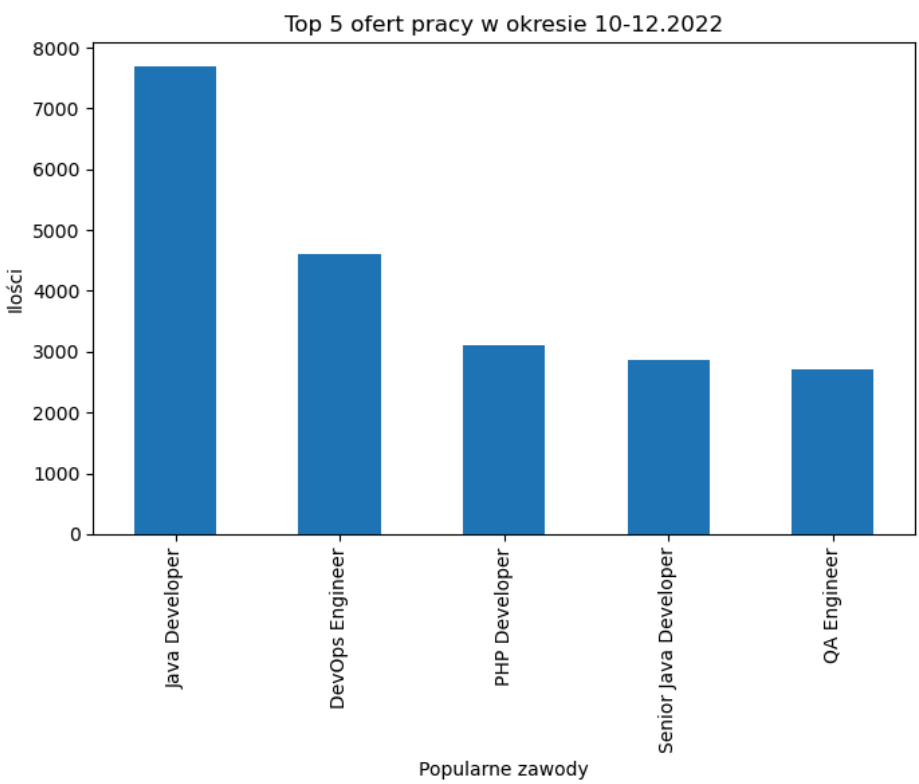
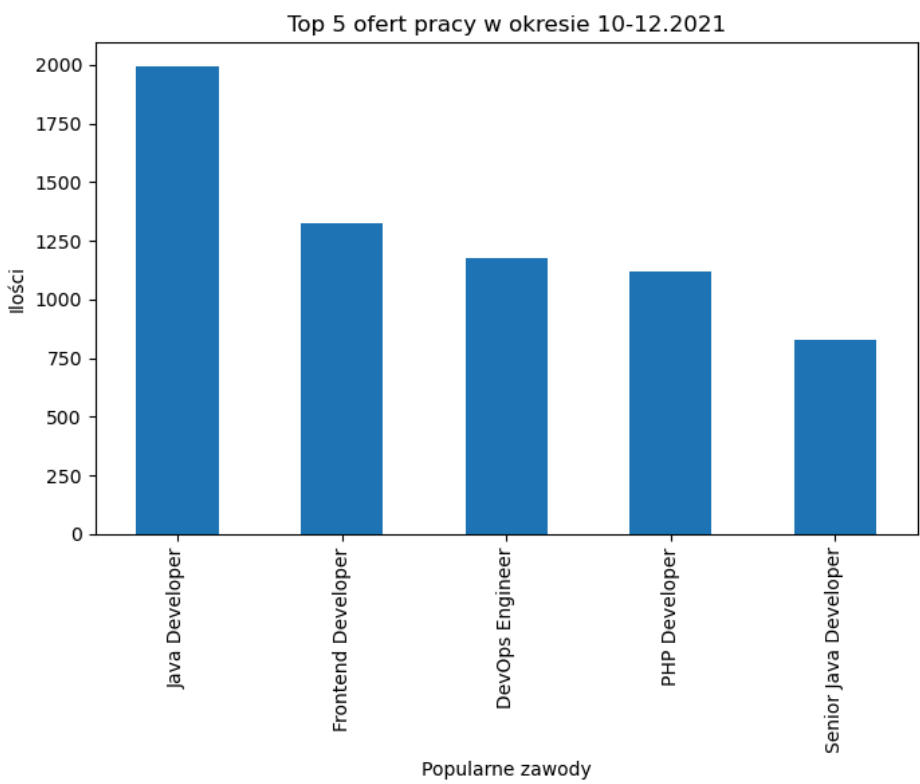
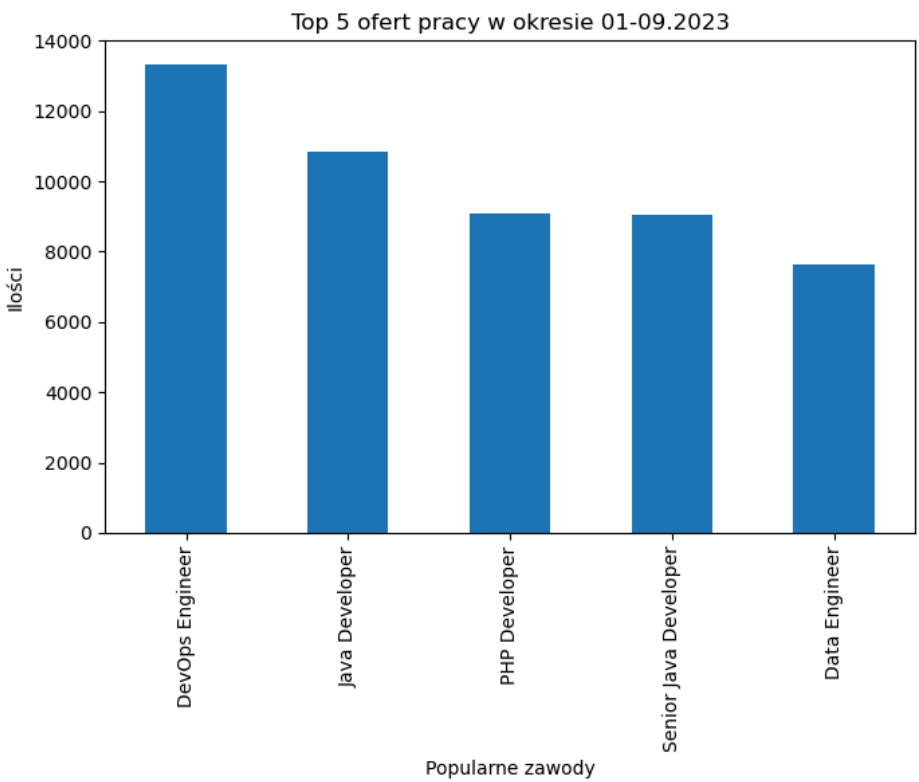
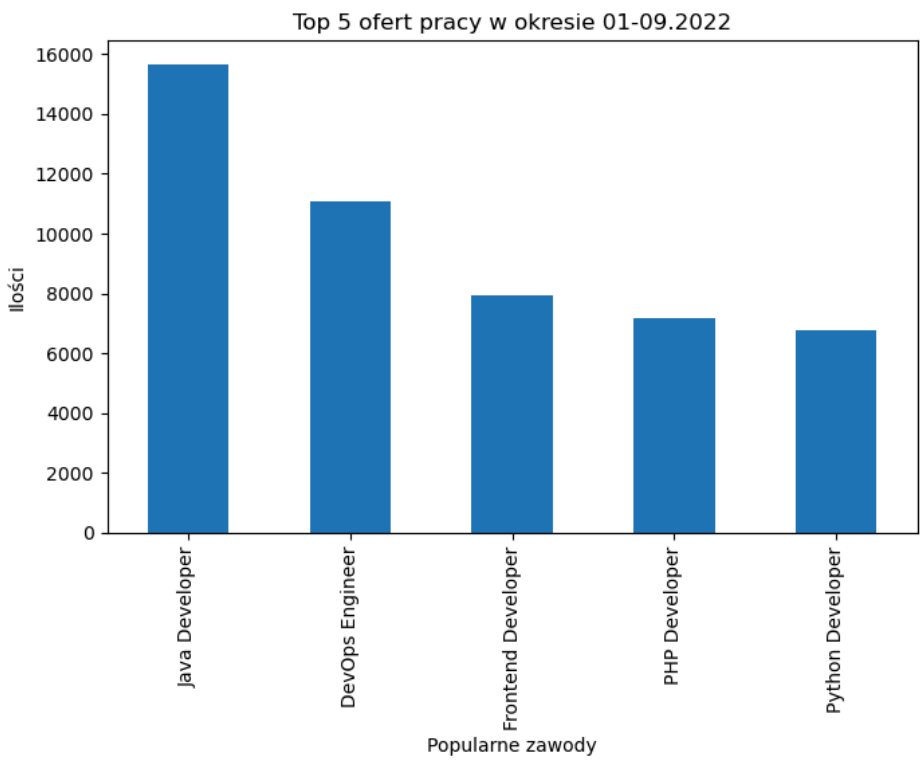
- Po wstępnych obserwacjach wynagrodzeń w ofertach pracy znaleziono wynagrodzenia w obcych walutach.
 - wszystkie waluty zagraniczne przekonwertowane na pln dla lepszego zrozumienia danych.
- Obserwacje minimalnego i maksymalnego wynagrodzenia - dokonano naprawy wartości w danych.
 - przytuszczalnie najmniejsze wynagrodzenia odnosiły się do stawek dziennych przez co zostały podniesione do wartości miesięcznej.
 - wartości max wynagrodzeń zostały przypuszczalnie zinterpretowane na wynagrodzeń rocznych i zostały poprawione na wynagrodzenia miesięczne.
- Obeserwacje po naprawie danych pokazują przedziały płacowe na bardzo zróżnicowanym poziomie.
 - W typach współpracy 'b2b' oraz 'permanent' znacząca ilość wartości odstających przekraczających maksymalne średnie wynagrodzenie.

4. Analiza trendów

1) Analiza archiwalnych ofert pracy:

- W okresie od Października do Grudnia 2021 roku opublikowano: **92908** ofert pracy.
- W okresie od Października do Grudnia 2022 roku opublikowano: **335067** ofert pracy.
- Porównując dane z roku 2022 a z 2021 widać wzrost publikacji ogłoszeń aż o: **260.64%**,
przez co można stwierdzić rozwój branży IT i zapotrzebowanie na nowych specjalistów.

- W okresie od Stycznia do Września 2022 roku opublikowano: **780772** ofert pracy.
- W okresie od Stycznia do Września 2023 roku opublikowano: **879133** ofert pracy.
- Co stanowi: **12.60%** wzrostu ofert pracy w danych okresach,
przez co można stwierdzić że zapotrzebowanie na wykwalifikowanych pracowników z roku na rok wzrasta.

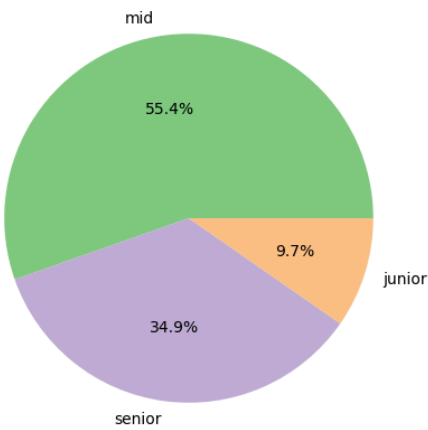


Obserwacje wstępne:

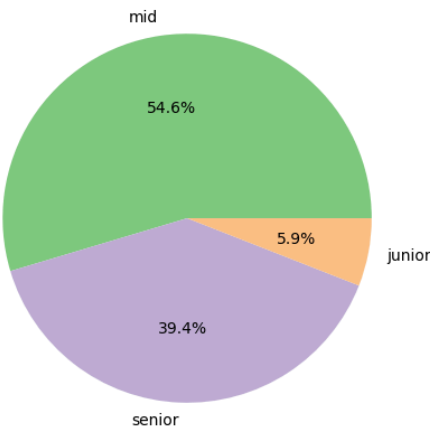
Porównując stanowiska pracy z ostatnich lat w danych okresach czasu stwierdzono że:

* Największe zapotrzebowanie było na stanowisko: **Java Developer**, **DevOps Enginner** oraz **PHP Developer**.

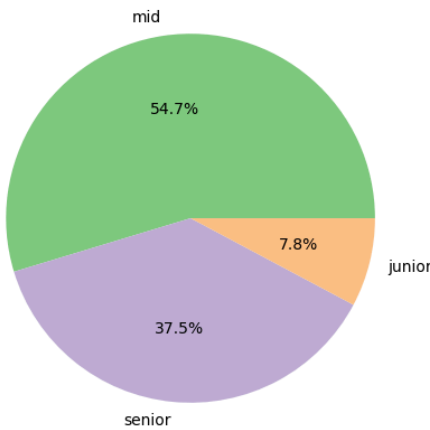
Poziom doświadczenia *10-12.2021*



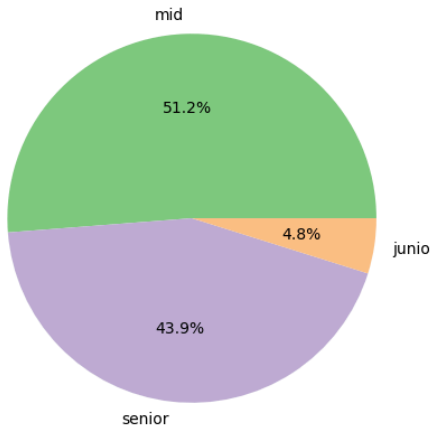
Poziom doświadczenia *10-12.2022*



Poziom doświadczenia *01-09.2022*



Poziom doświadczenia *01-09.2023*



Analiza archiwalnych ofert pracy z podziałem na poziomy doświadczenia:

- Z danych archiwalnych wynika że coraz więcej firm wymaga doświadczenia na poziomie *Senior*.
- W danych jest ponad połowa ofert pracy z wymaganym doświadczeniem na poziomie *Mid* .
- Niepokojącym trendem spadkowym możemy określić znaczne zmniejszenie zapotrzebowania na osoby z doświadczeniem na poziomie *Junior*.

Podsumowanie Analizy Danych

1. Świadomość Danych

- Źródła: Archiwalne dane ofert pracy z JostJoinIT
- Wielkość:
 - Rekordy: 2096717
 - Kolumny: 15

2. Podstawowe Statystyki

- Typy danych: bool, float, int, object.

3. Jakość Danych

- Brakujące wartości: Kolumny tekstowe(object) nie posiadają bakujących wartości, w kolumnach z płacami występują brakujące wartości.
- Zidentyfikowano błędy w kolumnach z wynagrodzeniami, ceny podawane w różnych walutach, prawdopodobnie w stawkach dniowych jak i rocznych zostały odpowiednio przygotowane do analizy.
- Duplikaty: Ponad połowa danych była duplikatami które zostały usunięte przed analizą.

4. Podsumowania i Wnioski

- Główne wnioski:
 - Dane poddane analizie zawierają : 33818 stanowisk pracy z 7676 firm.
 - Trzy najczęściej występujące stanowiska:
 - Java Developer
 - DevOps Enginer
 - PHP Developer
 - Z danych wynika że prawie wszystkie oferty pracy były zamieszczane w godzinach pracy: 7 - 16, reszta ofert była opublikowana w godzinach nocnych: 17 - 6, co może sugerować dodawanie ofert np. zagranicznych.
 - Dane przedstawiają trzy sposoby wykonywania pracy:
 - Praca zdalna: 74.93%
 - Praca hybrydowa: 22.38%
 - Praca biurowa: 2.69%
 - Biorąc pod uwagę okres który obejmuje wszystkie ogłoszenia można stwierdzić że najmniejsze zainteresowanie pracownikami biurowymi było spowodowane przez COVID-19.
 - Najwięcej ofert pracy jest w:
 - Warszawa 20.22%
 - Wrocław 11.52%
 - Kraków 11.44%
 - Aż 74.92% ofert pracy umożliwia rekrutację zdalną, a 35.26% z ofert umożliwia aplikowanie obcokrajowcom z Ukrainy.
 - Stwierdzono wzrost ilości ofert pracy, analizując dane z lat 2022-2023.
 - Stanowiska z wymaganym doświadczeniem na poziomie *Mid* cieszą się największą popularnością bo zajmują ponad 50% dostępnych ofert.
 - Przybywa ofert pracy z doświadczeniem na poziomie *Senior*.
 - Bardzo niepokojącym odkryciem jest że ilość ofert pracy z początkującym doświadczeniem na poziomie *Junior*, których w zbiorze danych jest zaledwie 6.3%. Po wykonaniu głębszej analizy zauważono spadek tych stanowisk o około 40%.
 - Przedziały płacowe sa na bardzo zróżnicowanym posioomie.
 - W typach współpracy b2b oraz permanent znacząca ilość wartości odstających przekracza maksymalne średnie wynagrodzenie.
- Następne kroki:
 - Zaleca się wykonanie porównania archiwalnych danych z nowymi w zakresie spadkowego trendu ofert pracy na rynku *Juniorskim*.
 - Zaleca się sprawdzenie trendów dla poziomów doświadczeń: *Senior* oraz *Mid*.