

Projekt Bazy Danych Firmy 1REMONT4U

Romana Żmuda(249706), Patryk Krukowski(249824), Adrian Kit(249746)

17 lutego 2021

Spis treści

1	Wstęp	1
2	Generowanie i wgrywanie danych	2
3	Analiza danych	4
3.1	Przygotuj wykres liczby oczekujących zleceń w czasie	5
3.2	Przygotuj wykres przedstawiający zmieniającą się liczbę czynnych pracowników każdego miesiąca.	6
3.3	Stwórz bilans finansowy firmy — czy jest w stanie się utrzymać?	6
3.4	Stwórz listę osób najdłużej czekających na remont	7
3.5	Własne pytania	8
3.5.1	Wyliczenie procentowego wkładu poszczególnych czynników na przycho- dy i wydatki	9
3.5.2	Stworzyć histogram czasu oczekiwania na remont, podać średni czas ocze- kiwania oraz medianę	9
3.5.3	Czas realizacji zleceń	10
3.5.4	Jaki sprzęt był najczęściej wykorzystywany?	11
4	Zakończenie	12

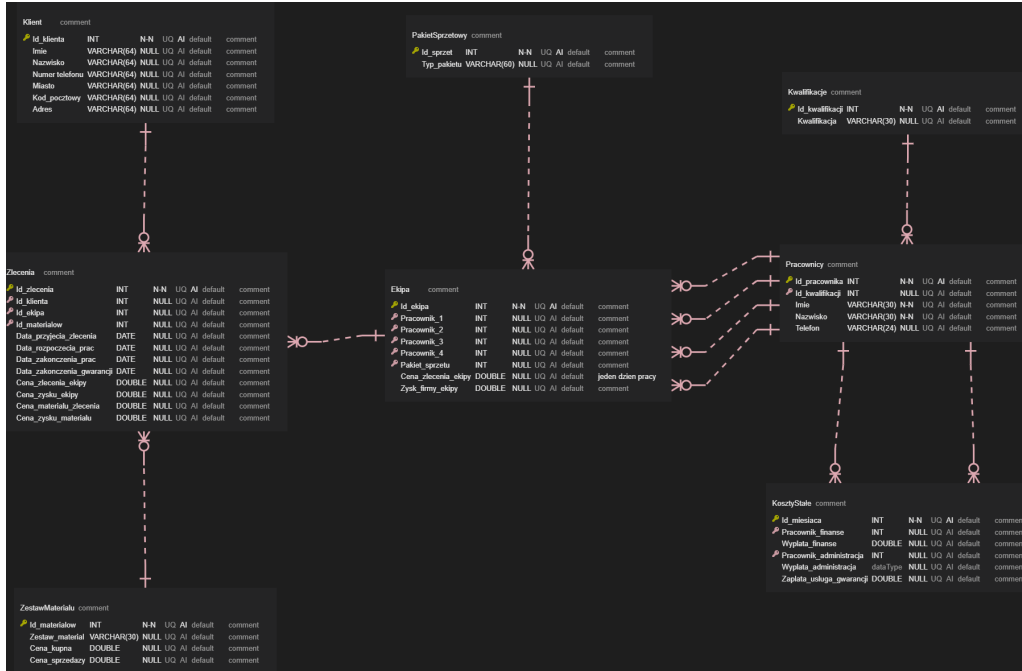
1 Wstęp

W tej części przedstawimy króciutki opis poszczególnych tabel w bazie danych, opowiemy o przygotowaniach i o tym jak powstała baza. Zaczniemy od opisu. Mamy łącznie 8 tabel:

- Klient - podstawowe dane o klientach firmy
- Zlecenia - informacje o zleceniach takie jak daty rozpoczęcia i zakończenia prac oraz informacje o wartościach danego zlecenia
- Pracownicy - podstawowe informacje o pracownikach
- Kwalifikacje - funkcja jaką pełni pracownik
- Ekipa - pracownicy są przydzieleni do ekip, w których pracują, na stałe. Znajdziemy tu także informacje o sprzęcie, jakiego używała ekipa oraz o finansach.

- PakietSprzetowy - sprzęt firmy podzieliliśmy w pakiety, które są przydzielone do ekip
- ZestawMaterialu - informacje o materiałach
- KosztyStale - wynagrodzenie pracowników biurowych i koszty serwisu gwarancyjnego

Wizualizacja bazy:



Analizę przeprowadzimy w R, dane na serwer wgraliśmy także przy użyciu R. Jeśli chodzi o generowanie to skorzystaliśmy z Excela i R.

2 Generowanie i wgrywanie danych

- Zaczęliśmy od generowania imion, problem rozbiliśmy na imiona męskie, żeńskie zrobiliśmy się analogicznie.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	SUMA	Nazwisko	Liczba		SUMA	Pierwsze imię	Liczba							
2	1	NOWAK	101036		1	JAN	1066985							
3	101037	KOWALSKI	68406		1066986	STANISLAW	935195							
4	155380	WISNIEWSKI	54343		2002181	PIOTR	907374							
5	204337	WOJCIK	48957		2809555	ANDRZEJ	743295							
6	252521	KOWALCZYK	48184		3552850	KRZYSZTOF	730968							
7	299467	KAMIŃSKI	46946		4283818	JÓZEF	687823							
8	344980	LEWANDOWSKI	45513		4971641	TOMASZ	577975							
9	389828	ZIELIŃSKI	44848		5549616	PAWEŁ	570756							
10	433500	WOŹNIAK	43672		6120372	MICHAŁ	554342							
11	477134	SZYMAŃSKI	43634		6674714	ADAM	481972							
12	519740	DĄBROWSKI	42606		7156686	TADEUSZ	477391							
13	556927	KOZŁOWSKI	37187		7634077	MARCIN	474163							
14	590918	MAZUR	33991		8108240	MAREK	455351							
15	624878	JANKOWSKI	33960		8563591	ERZYM	449514							
16	657545	KWIATKOWSKI	32667		9013105	GRZEGORZ	424637							
17	689869	WOJCIECHOWSKI	32324		9437742	JAKUB	413595							
18	721109	KRAWCZYK	31240		9851337	HENRYK	399039							
19	751720	KACZMAREK	30611		10250376	ZBIGNIEW	397285							
20	781808	PIOTROWSKI	30088		10647661	LUKASZ	396089							
21	810530	GRABOWSKI	28722		11043750	KAZIMIERZ	385332							
22	838165	ZAJĄC	27635		11427282	MATEUSZ	380055							
23	865518	PAWŁOWSKI	27353		11807337	WOJCIECH	378914							
24	892581	KRÓL	27063		12186251	RYSZARD	365786							
25	919292	MICHAŁSKI	26711		12552037	MARIAN	338470							
26	944497	WRÓBEL	25205		12890507	DARIUSZ	318534							
27	969339	WIECZOREK	24842		13209041	WŁADYSŁAW	314660							
28	994040	JABŁOWSKI	24701		13523701	MARIUSZ	311008							
29	1018285	NOWAKOWSKI	24245		13834709	ROBERT	284588							
30	1042488	MAJEWSKI	24203		14119297	FRANCISZEK	281496							

- Następnie wygenerowaliśmy daty potrzebne do realizacji zamówień i zgodnie z założeniami inne potrzebne nam kolumny.

BD.ILE.REK...		✕ ✓ <i>fx</i>		=LOS.ZAKR(DATA(2020;1;1);DATA(2021;2;9))	
	A	B	C	D	E
1	=LOS.ZAKR(DATA(2020;1;1);DATA(2021;2;9))				
2	2	LOS.ZAKR(dół; góra)	08.02.2021	10.03.2021	
3	2020-09-05		07.02.2021	04.03.2021	
4	2020-05-21		06.02.2021	04.03.2021	
5	2020-03-14		06.02.2021	09.02.2021	
6	2020-08-20		05.02.2021	24.02.2021	
7	2020-07-26		05.02.2021	27.02.2021	
8	2020-12-12		04.02.2021	19.02.2021	
9	2021-01-10		04.02.2021	11.02.2021	
10	2020-10-22		02.02.2021	13.02.2021	
11	2021-01-20		02.02.2021	02.03.2021	
12	2020-11-11		02.02.2021	08.02.2021	
13	2020-05-06		01.02.2021	13.02.2021	
14	2020-06-24		01.02.2021	15.02.2021	
15	2020-05-20		31.01.2021	09.02.2021	
16	2020-01-16		31.01.2021	02.02.2021	
17	2020-04-03		31.01.2021	14.02.2021	
18	2020-03-09		30.01.2021	18.02.2021	
19	2020-05-06		30.01.2021	07.02.2021	
20	2020-12-12		29.01.2021	09.02.2021	
21	2021-01-18		29.01.2021	11.02.2021	
22	2021-01-09		28.01.2021	09.02.2021	
23	2020-04-06		28.01.2021	13.02.2021	
24	2021-01-28		28.01.2021	17.02.2021	
25	2020-02-29		28.01.2021	10.02.2021	
26	2020-05-15		27.01.2021	24.02.2021	
27	2020-11-20		26.01.2021	16.02.2021	
28	2020-07-12		26.01.2021	06.02.2021	
29	2021-01-25		24.01.2021	04.02.2021	

✕ ✓ <i>fx</i>		=(LOS()*(31-2)+2)+C2	
B	C	D	
	Data przyjęcia	Data rozpoczęcia prac	
	08.02.2021	=(LOS()*(31-2)+2)+C2	
	07.02.2021	08.03.2021	

Nasza firma prowadzi działalność od 1 stycznia 2020r., w naszej bazie zlecenia przyjmowaliśmy do 9 lutego 2021r. Pracę musimy rozpocząć najpóźniej miesiąc (31 dni) od dnia przyjęcia zlecenia, zaś sam remont nie może trwać dłużej niż 14 dni.

- Gotową bazę przygotowaną w Excelu wgraliśmy do R za pomocą biblioteki `xlsx`. Kłopotem jaki napotkaliśmy było odpowiednie przyporządkowanie ekip do zlecenia tak, by w jednym

czasie ekipa nie robiła dwóch zleceń naraz.

```
begin <- as.character(dane$Data_roz poczenia_prac)
end <- as.character(dane$Data_zakonczenia_prac)
id_ekipy <- as.character(dane$Id_ekipa)

id_ekipy <- rep(0,1000)

for (k in 1:length(id_ekipy)){
  start <- begin[k]
  koniec <- end[k]

  dostepna <- rep(TRUE, 50)
  for (j in 1:k){
    if (
      (start <= begin[j] & koniec >= begin[j]) |
      (start <= end[j] & koniec >= end[j]) |
      (start >= begin[j] & koniec <= end[j])
    ){
      dostepna[id_ekipy[j]] <- FALSE
    }
  }
  print(paste(c("W kroku", k, "dostepne:", which(dostepna))))
  id_ekipy[k] <- sample(which(dostepna), 1)
}

id_ekipy
Id_ekipy<-data.frame(id_ekipy)
```

- Tak przygotowaną bazę wgraliśmy na serwer za pomocą biblioteki *RMariaDB*

```
#wgrywanie danych
install.packages("xlsx")
library(xlsx)

dane1<-read.xlsx2("C:/Users/roman/OneDrive/Pulpit/BAZA_projekt/Tabele_wypełnione.xlsx",
  sheetName = "ZestawMaterialu", header = TRUE)

#dbConnect z RMariaDB
install.packages("RMariaDB")
library(RMariaDB)
# to musisz zrobić, jak chcesz robić cokolwiek z serwerem
con_team4 <- dbConnect(RMariaDB::MariaDB(), host = "giniewicz.it",
  user = "team4", password = "t3@mAP@55", dbname="team4")
#wgrywanie danych do GOTOWYCH TABEL, podobna funkcja (READ) robi ze z sql do R
dbWriteTable(
  con_team4,
  "Zlecenia",
  dane, row.names=FALSE, append = TRUE)
```

3 Analiza danych

Baza została połączona z R. Właśnie w nim wykonamy naszą analizę. Podstawowe zadania do wykonania:

- Przygotuj wykres liczby oczekujących zleceń w czasie.
- Przygotuj wykres przedstawiający zmieniającą się liczbę czynnych pracowników każdego miesiąca.
- Stwórz bilans finansowy firmy — czy jest w stanie się utrzymać?

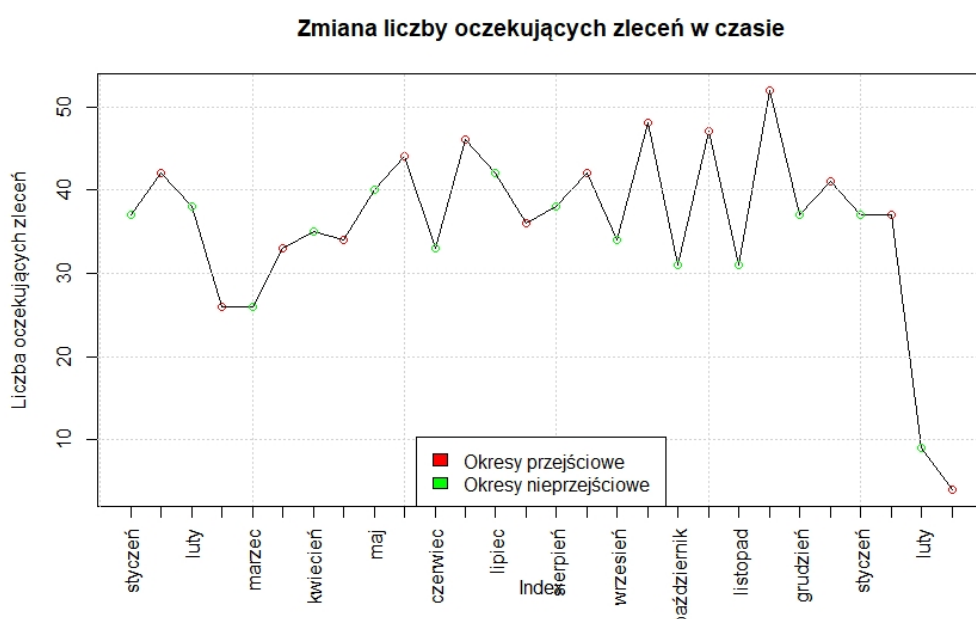
- Stwórz listę osób najdłużej czekających na remont.

Naszą analizę poszerzyliśmy o problemy:

- Stworzyć histogram czasu oczekiwania na remont, podać średni czas oczekiwania oraz medianę.
- Wyliczenie procentowego wkładu poszczególnych czynników na przychody i wydatki
- Czas realizacji zleceń
- Najczęściej wykorzystywany sprzęt

W kolejnych podsekcjach będziemy omawiać szczegółowo powyższe problemy.

3.1 Przygotuj wykres liczby oczekujących zleceń w czasie



Opis wykresu: Kolorem zielonym oznaczyliśmy punkty odpowiadające ilości zleceń rozpoczętych i zakończonych w i-tym miesiącu (okresy nieprzejściowe), natomiast czerwone odpowiadają zleceniom rozpoczętym w i-tym i zakończonym w i+1 miesiącu (okresy przejściowe).

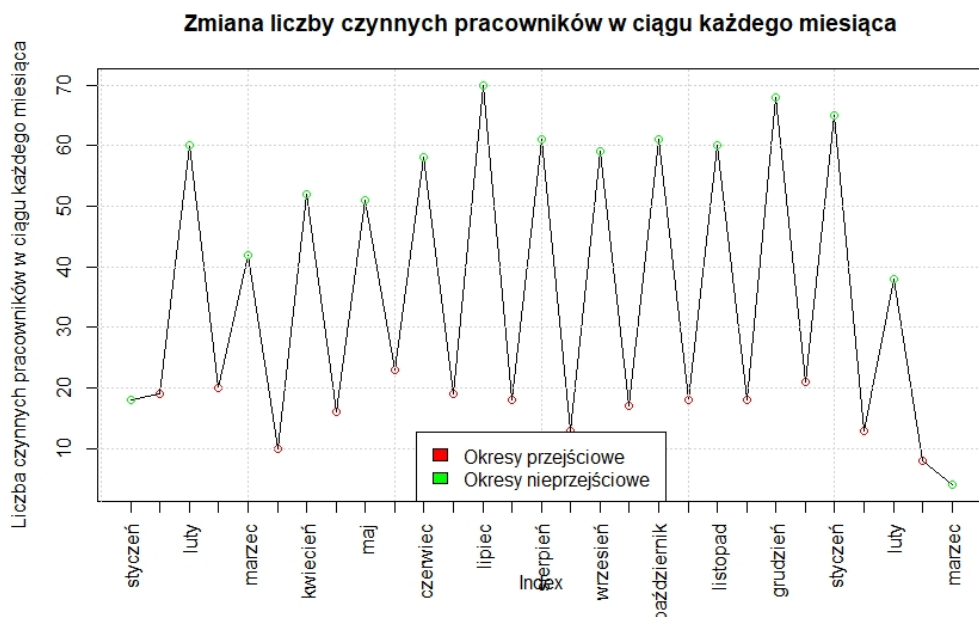
Obserwacje:

- Najwięcej zleceń było przyjmowanych w okresie od września do listopada. Ciekawym pomysłem jest sprawdzenie, jak bardzo jest to odzwierciedlone w dochodach. Stąd można wyciągnąć także wiele innych, cennych dla firmy informacji.
- Niewielka ilość zleceń w ostatnich dwóch miesiącach (jest to spowodowane "końcem" naszej bazy, w lutym zostało przyjęte ostatnie zlecenie)
- Można zauważyć, że więcej zleceń było przyjmowanych w okresach przejściowych

Te cenne informacje to:

- Firma powinna przygotować się na wzmożoną aktywność, na przykład zatrudnić więcej pracowników na ten okres, ostrzec hurtownie i próbować zbić cenę materiałów itp.
- W kalkulacjach zarząd firmy powinien uwzględnić kapitał zapasowy na wynagrodzenie dla pracowników w trakcie pozostałego okresu celem zachowania płynności finansowej.

3.2 Przygotuj wykres przedstawiający zmieniającą się liczbę czynnych pracowników każdego miesiąca.



Obserwacje i wnioski:

- Widać wyraźne skoki wartości, które na pierwszy rzut oka są losowe, jednak po chwili głębszej analizy zauważamy istotną zależność: Liczba czynnych pracowników maleje zawsze w pierwszej połowie miesiąca, tzn. od końca okresu nieprzejściowego do przejściowego i rośnie od okresu przejściowego do nieprzejściowego. To wyraźnie świadczy o większym zainteresowaniu remontami w drugiej połowie miesiąca, co potwierdza rezultaty z zadania pierwszego
- Pomimo powyższego wniosku, liczba czynnych pracowników nie wzrosła wyraźnie w okresie wzmożonej aktywności (na wykresie z zadania pierwszego była wzmożona aktywność w miesiącach wrzesień-listopad, tutaj brak takiej anomalii)
- Najwięcej czynnych pracowników było w pierwszej połowie lipca, nie pokrywa się to jednak z rezultatami zadania pierwszego.

3.3 Stwórz bilans finansowy firmy — czy jest w stanie się utrzymać?

W naszej firmie bilans wyliczamy w następujący sposób: Jest to kwota, którą zarobiliśmy na zleceniu wykonywanym przez daną ekipę oraz zysk z materiałów pomniejszona o koszty stałe i cenę materiału, jaką zapłaciliśmy hurtowni.

```
> zapytanie_3_koszty_stale <- dbGetQuery(db_con, "
+                               SELECT
+                               SUM(Wyplata_finanse) AS S1, SUM(Wyplata_administracja) AS
+                               SUM(Zaplata_usluga_gwarancji) AS S3
+                               FROM KosztyStale;
+                               ")
> koszty_stale_bilans <- sum(zapytanie_3_koszty_stale)#sumujemy wszystkie koszty stałe
> materialy_bilans <- as.double(dbGetQuery(db_con, "
```

```

+           SELECT SUM(Cena_kupna)
+           FROM ZestawMaterialu
+           ")
> bilans <- sum(zlecenia$Cena_zysku_ekipy)+sum(zlecenia$Cena_materialu_zlecenia)-koszt
> bilans

```

```
[1] 2521746
```

Okazuje się, że nasza firma jest w stanie się utrzymać. Przy wyliczaniu bilansu zaciekał nas wpływ poszczególnych czynników na dochód i koszty. Mówiąc konkretniej, zadaliśmy sobie pytanie, czy zysk z materiałów ma jakikolwiek wpływ na generowany przez firmę dochód. Z drugiej strony, co generuje największe koszty? Dodaliśmy więc to pytanie do puli własnych problemów do rozwiązania.

3.4 Stwórz listę osób najdłużej czekających na remont

Korzystamy z wbudowanej funkcji DATEDIFF pozwalającej obliczyć nam czas oczekiwania klientów na rozpoczęcie prac.

```

> zapytanie_4 <- dbGetQuery(db_con, "SELECT Imie, Nazwisko, DATEDIFF(Data_rozpoczecia_
+           FROM Klient
+           INNER JOIN Zlecenia
+           ON Zlecenia.Id_klienta = Klient.Id_klienta
+           ORDER BY Czas DESC
+           LIMIT 50
+           ")
> zapytanie_4

```

	Imie	Nazwisko	Czas
1	ANDRZEJ	WORONOWICZ	30
2	JÓZEF	WAWROCKI	30
3	MATEUSZ	URBAŃSKI	30
4	SZYMON	SYLDATK	30
5	EWA	SWÓDROWSKA	30
6	DANIELA	STAWARZ	30
7	DANUTA	SÓMA	30
8	AUGUSTYNA	SABURA	30
9	JÓZEF	ROGAŁA	30
10	JÓZEF	RADZISZEWSKI	30
11	REMIGIUSZ	PAŃUCHA	30
12	STANISŁAW	NADOLSKI	30
13	ANGELIKA	MICHALAK	30
14	RAFAŁ	MATUSZYK	30
15	JAN	KOZŁOWSKI	30
16	ŁUKASZ	KOROPETS	30
17	MAJA	KORNET	30
18	RENATA	KOPCYCH	30
19	JAN	KONDERAK	30
20	DARIUSZ	KALININ	30

21	JOANNA	J?DRYSIAK	30
22	KRZYSZTOF	JOD?OWSKI	30
23	BEATA	JE?EWSKA	30
24	MIROS?AW	JANUSZEWSKI	30
25	BARTOSZ	JANASZ	30
26	BERNARD	HUNTSCHA	30
27	FLORIAN	GÓRKA	30
28	MARCIN	GAJER	30
29	JAN	GAJAK	30
30	TOMASZ	DZIEKIEWICZ	30
31	STANIS?AW	DAWID	30
32	MICHALINA	BRYCHCY	30
33	OLEKSII	BOJANOWSKI	30
34	JERZY	BOGACKI	30
35	PIOTR	BIEGA?SKI	30
36	?UKASZ	BARTODZIEJ	30
37	KRYSTYNA	ANTCZAK	30
38	KRZYSZTOF	ADAMCZYK	30
39	MIROS?AW	SZCZERBACIUK	29
40	BERNARD	SYMULEWICZ	29
41	EUGENIUSZ	S?OMA	29
42	BOGUMI?A	ROGALA	29
43	ZOFIA	RADZIK	29
44	KINGA	PUCHALSKA	29
45	BLANKA	POTOCKA	29
46	JAN	POKORNICKI	29
47	BOGUMI?A	PODOLNA	29
48	JAN	PIECHOWSKI	29
49	ANTONI	PACHALY	29
50	NATALIA	ORSZULAK	29

Rozwiązując to zadanie zaciekało nas, ile wynosi średni czas oczekiwania na remont w naszej firmie oraz jak wypadamy na tle rzeczywistych firm, celem weryfikacji, czy mimo losowo generowanych danych nasza firma jest choć w niewielkim stopniu zbliżona do tych istniejących naprawdę. W tym celu obliczyliśmy średni czas oczekiwania na remont i porównaliśmy go z danymi znalezionymi w internecie. Rezultaty przedstawimy w następnej części dotyczącej własnych pytań i analizy.

3.5 Własne pytania

Naszą analizę postanowiliśmy poszerzyć o następujące problemy:

- Kontynuacja zadania trzeciego: Wyliczenie procentowego wkładu poszczególnych czynników na przychody i wydatki
- Kontynuacja zadania czwartego: Stworzyć histogram czasu oczekiwania na remont, podać średni czas oczekiwania oraz medianę.
- Czas realizacji zleceń
- Najczęściej wykorzystywany sprzęt

3.5.1 Wyliczenie procentowego wkładu poszczególnych czynników na przychody i wydatki

```
> zysk <- sum(zlecenia$Cena_zysku_ekipy) +  
+          sum(zlecenia$Cena_materialu_zlecenia)  
> (sum(zlecenia$Cena_zysku_ekipy)/zysk)*100
```

```
[1] 83.33719
```

```
>
```

Zysk z remontów wyniósł 83% , zatem zysk z materiałów to około 17% . To o wiele więcej niż przypuszczaliśmy. Analogiczna analiza dla wydatków, zaprezentowana w formie wykresu kołowego.

```
> pensja_stacjonarnych <- sum(koszty$Wyplata_finance)+sum(koszty$Wyplata_administracja)  
> koszt_serwis_gwarancyjny <- sum(koszty$Zaplata_usluga_gwarancji)  
> pensja_robotnikow <- sum(zlecenia$Cena_zlecenia_ekipy) - sum(zlecenia$Cena_zysku_eki)  
> wydatki <- pensja_stacjonarnych+koszt_serwis_gwarancyjny+pensja_robotnikow  
> x<-pensja_robotnikow/wydatki  
> x
```

```
[1] 0.9752001
```

Pensja pracowników fizycznych stanowi 97.5% wszystkich wydatków firmy.



Widzimy miażdżącą przewagę pensji pracowników fizycznych.

3.5.2 Stworzyć histogram czasu oczekiwania na remont, podać średni czas oczekiwania oraz medianę

```
> zapytanie_4.75 <- dbGetQuery(db_con, "SELECT Imie, Nazwisko, DATEDIFF(Data_rozpoczec  
+                                     FROM Klient  
+                                     INNER JOIN Zlecenia
```

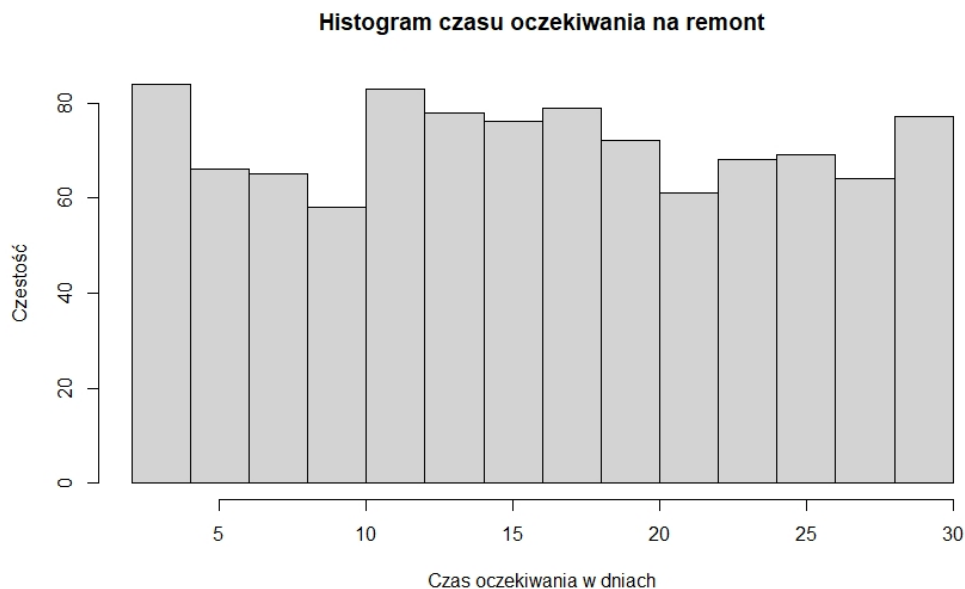
```

+           ON Zlecenia.Id_klienta = Klient.Id_klienta
+           ORDER BY Czas DESC
+       ")
> hist(zapytanie_4.75$Czas, main='Histogram czasu oczekiwania na remont', xlab = 'Czas
+       ylab='Częstość')
> summary(zapytanie_4.75)

```

Imie	Nazwisko	Czas
Length:1000	Length:1000	Min. : 2.00
Class :character	Class :character	1st Qu.:10.00
Mode :character	Mode :character	Median :16.00
		Mean :16.31
		3rd Qu.:23.00
		Max. :30.00

```
>
```



Widzimy, że średni i najczęstszy czas oczekiwania wynosi 16 dni. Próbowaliśmy porównać naszą firmę z tymi rzeczywistymi, jednak w internecie brak konkretnych informacji na ten temat. Doczytaliśmy, że realizację można przyspieszyć dopłacając dodatkowo. Jest to pomysł który powinien być znaleźć się w ofercie naszej firmy.

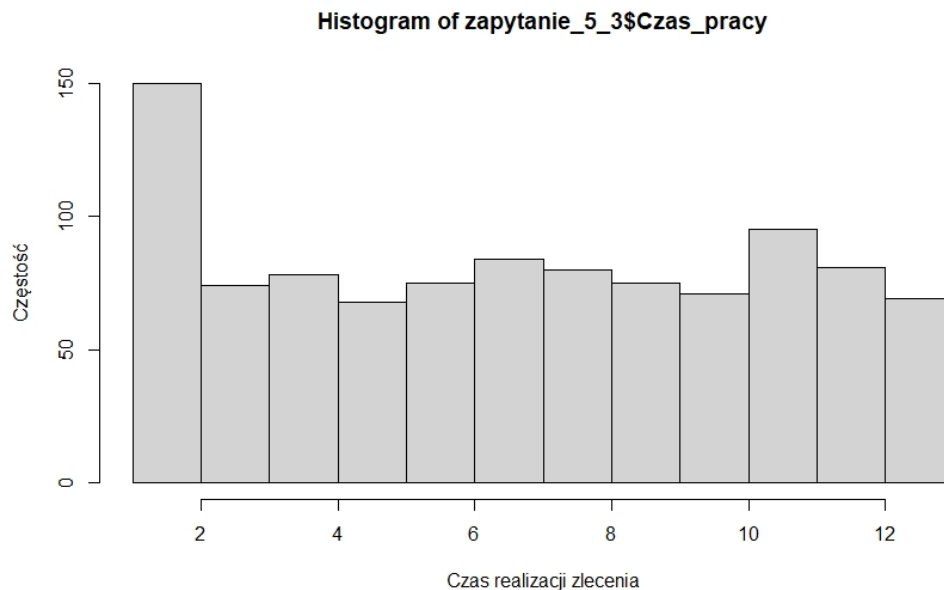
3.5.3 Czas realizacji zleceń

```

> zapytanie_5_3 <- dbGetQuery(db_con, "
+       SELECT Id_ekipa, Id_zlecenia,
+       DATEDIFF(Data_zakonczenia_prac, Data_rozpoczecia_prac) A
+       FROM Zlecenia
+       GROUP BY Id_ekipa, Id_zlecenia
+       ORDER BY Czas_pracy DESC;
+   ")
> hist(zapytanie_5_3$Czas_pracy, xlab="Czas realizacji zlecenia", ylab="Częstość")
> summary(zapytanie_5_3$Czas_pracy)

```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.000	4.000	7.000	7.074	10.000	13.000



Najdłuższe zlecenia były wykonywane 13 dni, natomiast najkrótsze tylko 1 dzień. Zaskakująco wysoko liczba zleceń jednodniowych. Przeważają zlecenia trwające tydzień czasu.

3.5.4 Jaki sprzęt był najczęściej wykorzystywany?

```
> zapytanie_5_4 <- dbGetQuery(db_con, "
+      SELECT Pakiet_sprzetu, COUNT(Pakiet_sprzetu) AS
+      Ilosc_uzyc
+      FROM Ekipa
+      INNER JOIN Zlecenia
+      ON Ekipa.Id_ekipa = Zlecenia.Id_ekipa
+      GROUP BY Pakiet_sprzetu
+      ORDER BY Ilosc_uzyc DESC
+      LIMIT 5;
+    ")
```

```
> zapytanie_5_4
```

	Pakiet_sprzetu	Ilosc_uzyc
1	33	38
2	32	32
3	42	27
4	13	24
5	31	24

```
>
>
```

Pakiet numer 33 został użyty najwięcej razy, bo aż 38. Niestety przez niedokładne opisanie bazy danych nie jesteśmy w stanie powiedzieć, co wchodziło w jego skład, przez do odbieramy sobie możliwość wyciągnięcia pożądaných wniosków.

4 Zakończenie

Najwięcej problemów sprawiły nam dwa pierwsze zadania z analizy oraz brak doświadczenia w generowaniu zmiennych. Przy wgrywaniu problemem było zachowanie odpowiednich wartości funkcji jako zmiennych w danym poleceniu, tak, by dane się dopisywały, a nie tworzyły nową tabelę. Przy wgrywaniu należało pamiętać o zachowaniu odpowiadającego nazewnictwa kolumn SQL z Excelem.