Spis treści

[Obserwacje wynikające z Eksploracyjnej Analizy Danych 3](#_Toc134353465)

[Kiedy przysyłanych jest najwięcej dokumentów? 4](#_Toc134353466)

[Duration X dzień X godzina 6](#_Toc134353467)

[Duration X dzień X tydzień 7](#_Toc134353468)

[Jakie są osiągane czasy przetwarzania? (zmienna ProcesingDuration) 28](#_Toc134353469)

[Eksploracyjna Analiza Danych – potok\_1-1 33](#_Toc134353470)

[**Zbiór danych** - typy danych 33](#_Toc134353471)

[**Zbiór danych** - generalne statystyki 33](#_Toc134353472)

[**Zbiór danych** - ostrzeżenia 33](#_Toc134353473)

[**Zmienna - InputPageCount** - charakterystyki zmiennej 34](#_Toc134353474)

[**Zmienna - MessagesInQueue** - charakterystyki zmiennej 37](#_Toc134353475)

[**Zmienna - Retries** - charakterystyki zmiennej 39](#_Toc134353476)

[**Zmienna - ProcessingDuration\_sec** - charakterystyki zmiennej 40](#_Toc134353477)

[**Zmienna - Duration\_sec** - charakterystyki zmiennej 42](#_Toc134353478)

[**Zmienna - StartDateTimeUtc** - charakterystyki zmiennej 44](#_Toc134353479)

[**Tablice korelacji** – korelacja między zmiennymi 45](#_Toc134353480)

[**Tablice korelacji** – korelacja między brakami danych pomiędzy zmiennymi 46](#_Toc134353481)

[Eksploracyjna Analiza Danych – potok\_1-2 47](#_Toc134353482)

[Eksploracyjna Analiza Danych – potok\_3-1 47](#_Toc134353483)

[Eksploracyjna Analiza Danych – potok\_3-2 47](#_Toc134353484)

[**Zbiór danych** - typy danych 47](#_Toc134353485)

[**Zbiór danych** - generalne statystyki 47](#_Toc134353486)

[**Zbiór danych** - ostrzeżenia 47](#_Toc134353487)

[**Zmienna - InputPageCount** - charakterystyki zmiennej 48](#_Toc134353488)

[**Zmienna - MessagesInQueue** - charakterystyki zmiennej 50](#_Toc134353489)

[**Zmienna - Retries** - charakterystyki zmiennej 52](#_Toc134353490)

[**Zmienna - ProcessingDuration\_sec** - charakterystyki zmiennej 54](#_Toc134353491)

[**Zmienna - Duration\_sec** - charakterystyki zmiennej 56](#_Toc134353492)

[**Zmienna - StartDateTimeUtc** - charakterystyki zmiennej 58](#_Toc134353493)

[**Tablice korelacji** – korelacja między zmiennymi 59](#_Toc134353494)

[**Tablice korelacji** – korelacja między brakami danych pomiędzy zmiennymi 60](#_Toc134353495)

Obserwacje wynikające z Eksploracyjnej Analizy Danych

* Dla celów wnioskowania na temat statystyk całego Zbioru Danych odrzucam zdarzenia, gdy pole *ID\_Document* jest wartością pustą. Innymi słowy: w celu wyliczenia Statystyk Globalnych odrzucam te zdarzenia z dalszej analizy. Dokumentów bez przypisanego Potoku w Zbiorze Danych jest 1.2% . Nie jest mi znana interpretacja biznesowa zdarzeń, gdy *ID\_Document* przyjmuje wartość pustą.
* Nie wszystkie potoki były funkcyjne każdego dnia (przykład: potok\_3-2). Być może wynikało to z awarii na Środowisku Produkcyjnym lub otrzymany Zbiór Danych zawiera informacje o zdarzeniach ze Środowiska Deweloperskiego, w tym Zdarzenia związane z okresami Testowymi. Ze względu na powyższe zaleca się analizę każdego z Potoków osobno. Zaleca się wyciąganie wniosków oraz tworzenie rekomendacji w oparciu o statyki dla każdego z Potoków osobno.
* Wszystkie Zmienne numeryczne opisujące Potoki: *Duration*, *ProcessingDuration*, *InputPageCount*, *Retries*, ze względu na swój charakter biznesowy – procesowanie faktur w oparciu o rozwiązanie OCR (ang. *optical character recognition*), charakteryzują się rozkładami silnie prawo skośnymi. Ze względu na ograniczone możliwości procesowania, zdarzenia mogą być kolejkowane. Niemniej najczęściej Klienci procesują pojedyncze faktury.
* Historia zdarzeń potok\_2-1 (N: 1193) oraz potok\_2-2 (N: 497) dotyczy wyłącznie jednego dnia – tzn. poniedziałku 8 tygodnia 2023 roku. Wydaje się niezasadnym wnioskowanie na podstawie danego potoku w oparciu o historię zdarzeń z wyłącznie jednego dnia. Zdarzenia związane z powyższymi potokami nie zostały uwzględnione w dalszych analizach.

# Kiedy przysyłanych jest najwięcej dokumentów?

Rozkład zmiennej będącej flagą czasu wszystkich zdarzeń prezentuje się następująco:  
  
Obraz zawierający tekst, zrzut ekranu, Czcionka, diagram

Opis wygenerowany automatycznie

Na podstawie powyższego histogramu ciężko jednoznacznie stwierdzić kiedy i dlaczego procesowanych jest więcej dokumentów.

Z dużą dozą pewności założyć można, że więcej dokumentów procesowanych jest w tzw. godzinach biznesowych, w tzw. dni pracujące. Mniej więcej od 8:00 rano do 17:00 popołudniu.

Faktycznie histogram ze skumulowanej liczebności zdarzeń w podziale na dni tygodnia potwierdza intuicyjną tezę, że w łikendy procesowanych jest mniej faktur.

Obraz zawierający zrzut ekranu, diagram, tekst, Prostokąt

Opis wygenerowany automatycznie

Z kolei w ujęciu tygodniowym dostrzec można sporą zmienność w liczebności zdarzeń dla każdego z potoków. Niemniej w łikendy wciąż procesuje się znaczne liczby dokumentów.

Obraz zawierający zrzut ekranu, linia, Wykres

Opis wygenerowany automatycznie

[Histogram skumulowanych liczebności zdarzeń w podziale na dni oraz tygodnie](https://github.com/KonuTech/yet-another-analysis/blob/main/00_hisotgram_stacked_days_x_weeks.png)

Warto zauważyć, że na przestrzeni dwóch miesięcy zmniejszyła się liczba zdarzeń związanych z  
potok\_1-1 na rzecz zdarzeń związanych z potok\_2-1.

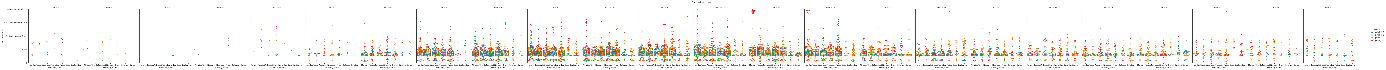
Większy popyt na usługi procesowania faktur w czasie tzw. godzin biznesowych powoduje, że część faktur trafia w tzw. kolejkę. Tym samym w dalszej części analiz do wymiarów czasu: godzin, dnia, tygodnia, miesięcy dodam zmienne *Duration* oraz *ProcessingDuration*.

Tzw. wykres roju pozwalają ocenić alokację zdarzeń w podzielę na kategorie.

## Duration X dzień X godzina

Poniższy wykres przedstawia próbkę 10% zdarzeń (około 100 tys. zdarzeń) w podziale na potok, dni tygodnia i godziny. Wykres pozwala na wstępną ocenę popytu na Usługi z uwzględnieniem wymiaru czasu.

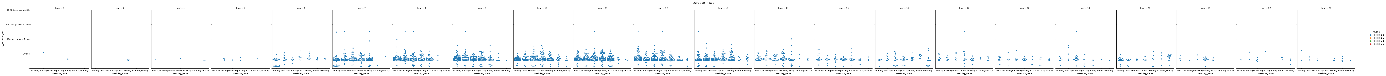
**Wszystkie potoki:**



Większy wykres:  
[Wykres roju dla wszystkich potoków – Duration w sek. dla dla godziny i tygodnia](https://github.com/KonuTech/yet-another-analysis/blob/main/01_swarm_duration_x_days_x_hours.png)

Z punktu widzenia każdego z potoków z osobna, wykresy rojowe przedstawiają podobny popyt na usługi w interakcji pomiędzy dniami tygodnia a godzinami.

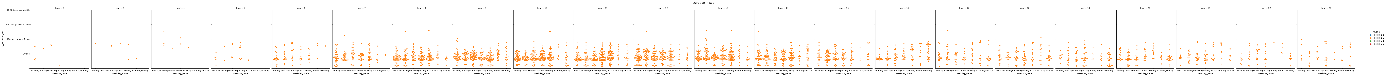
**potok\_1-1:**



Większy wykres:

[Wykres roju potok\_1-1 – Duration w sek. dla godziny i tygodnia](https://github.com/KonuTech/yet-another-analysis/blob/main/01_swarm_potok_1_1_duration_x_days_x_hours.png)

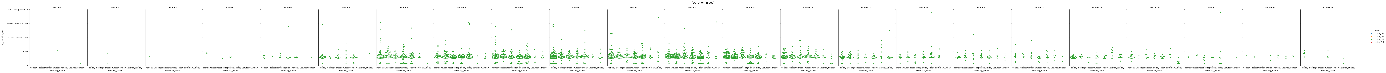
**potok\_1-2:**



Większy wykres:

[Wykres roju potok\_1-2 – Duration w sek. dla godziny i tygodnia](https://github.com/KonuTech/yet-another-analysis/blob/main/01_swarm_potok_1_2_duration_x_days_x_hours.png)

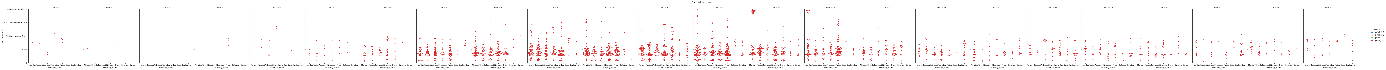
**potok\_3-1:**



Większy wykres:

[Wykres roju potok\_3-1 – Duration w sek. dla godziny i tygodnia](https://github.com/KonuTech/yet-another-analysis/blob/main/01_swarm_potok_3_1_duration_x_days_x_hours.png)

**potok\_3-2:**



Większy wykres:

[Wykres roju potok\_3-2 – Duration w sek. dla godziny i tygodnia](https://github.com/KonuTech/yet-another-analysis/blob/main/01_swarm_potok_3_2_duration_x_days_x_hours.png)

**Spostrzeżenia:**

Z perspektywy wszystkich Potoków, najwięcej dokumentów procesowanych jest:

* między godziną 7:00 a 15:00
* wciąż sporo faktur procesowanych jest między 5:00 a 6:00 oraz między 15:00 a 20:00
* znacznie mniej faktur procesowanych jest między 20:00 a 24:00
* istotnie niewiele faktur procesowanych jest między 24:00 a 6:00.

## Duration X dzień X tydzień

Poniższy wykres przedstawia 10% zdarzeń (około 100 tys. zdarzeń) w podziale na potok, dni tygodnia i tygodnie. Wykres pozwala na wstępną ocenę popytu na Usługi z uwzględnieniem wymiaru czasu.

[Wykres roju dla wszystkich potoków – Duration w sek. dla dnia i tygodnia](https://github.com/KonuTech/yet-another-analysis/blob/main/02_swarm_duration_x_days_x_weeks.png)

Wykresy rojowe dla każdego z Potoków przedstawiają podobne zaangażowanie zdarzeń w interakcji pomiędzy dniem tygodnia a godziną.

**potok\_1-1:**

Obraz zawierający tekst, linia, Wykres, Czcionka

Opis wygenerowany automatycznie

Większy wykres:

[Wykres roju potok\_1-1 – Duration w sek. dla dnia i tygodnia](https://github.com/KonuTech/yet-another-analysis/blob/main/02_swarm_potok_1_1_duration_x_days_x_weeks.png)

**potok\_1-2:**

Obraz zawierający tekst, linia, Wykres, Czcionka

Opis wygenerowany automatycznie

Większy wykres:

[Wykres roju potok\_1-2 – Duration w sek. dla dnia i tygodnia](https://github.com/KonuTech/yet-another-analysis/blob/main/02_swarm_potok_1_2_duration_x_days_x_weeks.png)

**potok\_3-1:**

Obraz zawierający tekst, zrzut ekranu, linia, Wykres

Opis wygenerowany automatycznie

Większy wykres:

[Wykres roju potok\_3-1 – Duration w sek. dla dnia i tygodnia](https://github.com/KonuTech/yet-another-analysis/blob/main/02_swarm_potok_2_1_duration_x_days_x_weeks.png)

**potok\_3-2:**

Obraz zawierający tekst, linia, Czcionka, Wykres

Opis wygenerowany automatycznie

Większy wykres:

[Wykres roju potok\_3-2 – Duration w sek. dla dnia i tygodnia](https://github.com/KonuTech/yet-another-analysis/blob/main/02_swarm_potok_2_2_duration_x_days_x_weeks.png)

Spostrzeżenia:

* We wtorek i środę 8 tygodnia 2023 r. wszystkie potoki były wyłączone lub doznały awarii.
* Między tygodniem 5 a 11 2023 r. jedynie potok\_1-1 oraz potok\_1-2 reprezentowały stabilną historię zdarzeń.
* Być może w tygodniu 5 i 6 2023 r. potok\_1-1 doświadczył awarii. *Duration* dla potoku\_1-1 było w tygodniu 5 wyraźnie wyższe aniżeli w tygodniach 7-11 - kiedy to potok ten funkcjonował stabilnie. W poniedziałek, w tygodniu 6 wiele procesów potoku\_1-1 charakteryzowało *Duration* nawet na poziomie 3h.
* Warto zauważyć, że:
  + ze względu na występowanie nietypowych zdarzeń średnia *Duration* dla całego zbioru danych to 25 sekundy.
  + natomiast historyczna mediana *Duration* wynosi poniżej 4 sekund
  + z kolei pierwszy kwartyl wynosi poniżej 3 sekund
  + a 3 kwartyl wynosi 7 sekund.

Jako, że każde unikatowe ID\_Document może być przetwarzane wiele razy, to do oceny kiedy przesyłanych jest najwięcej dokumentów użyję agregatu pola ID\_Process.

Na pytanie kiedy przysyłanych jest najwięcej dokumentów mogą odpowiedzieć nam m. in. trzy miary: średnia liczba procesów, średnie *Duration procesów* oraz średnie *MessagesInQueue* procesu - dla każdej godziny. Uwzględniając liczbę miesięcy i dni.

Według poniższego wykresu liniowego, najwięcej dokumentów przesyłanych jest między 6:00 a 15:00. Z „godzinami szczytu” między 8:00 a 13:00. Warto zwrócić uwagę na obecność grupy Klientów, którzy procesują faktury po typowym godzinach biurowych.

Obraz zawierający tekst, diagram, Wykres, linia

Opis wygenerowany automatycznie

Suma zmiennej *Duration* transformowanej Log10 w sekundach potwierdza rozkład godzin, w których usługi charakteryzuje największe obciążenie.

Obraz zawierający diagram, Wykres, tekst, linia

Opis wygenerowany automatycznie

Bardzo podobne rozkłady prezentuje zmienna Suma z MessagesInQueue transformowana Log10.

Obraz zawierający diagram, tekst, Wykres, linia

Opis wygenerowany automatycznie

Liczebność *ID\_Processing* wg tygodnia, dnia tygodnia i godziny:

Obraz zawierający tekst, zrzut ekranu, Wykres, linia

Opis wygenerowany automatycznie

Liczebność *ID\_Processing* wg potoku, tygodnia, dnia tygodnia i godziny:

Obraz zawierający tekst, zrzut ekranu, Czcionka, Wykres

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, Wykres, Czcionka

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, Wykres, linia

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, Wykres, linia

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, Wykres

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, Wykres, linia

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, Wykres, Czcionka

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, Czcionka, linia

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, Czcionka, Wykres

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, Wykres, Wielobarwność

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, Wykres, linia

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, Wykres, linia

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, Wykres, linia

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, Wykres, linia

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, Wykres

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, Wykres, linia

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, Wykres, linia

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, Wykres, linia

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, Wykres, linia

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, Wykres, diagram

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, linia, Wykres

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, Wykres, linia

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, Wykres, diagram

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, diagram, linia

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, Wykres, linia

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, Czcionka, Wykres

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, Czcionka, linia

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, Wykres, linia

Opis wygenerowany automatycznie

Obraz zawierający tekst, zrzut ekranu, Wykres, linia

Opis wygenerowany automatycznie

# Jakie są osiągane czasy przetwarzania? (zmienna ProcesingDuration)

W każdym z potoków wystąpiły zdarzenia nazywane zdarzeniami„ekstremalnymi”, „odstającymi”.

Czasy przetwarzania były rozmaite dla wszystkich potoków.

Obraz zawierający tekst, zrzut ekranu, diagram, Wykres

Opis wygenerowany automatycznie

Zdarzenia odstające występowały każdego dnia tygodnia

Obraz zawierający tekst, zrzut ekranu, diagram, Wykres

Opis wygenerowany automatycznie

Zdarzenia odstające występowały zarówno w typowych godzinach biurowych jak i poza typowymi godzinami biurowymi, czy w łikendy.

potok\_1-1 (próbka 10%)

Obraz zawierający statek wodny, panorama

Opis wygenerowany automatycznie przy średnim poziomie pewności

potok\_1-2 (próbka 10%)

Obraz zawierający statek, statek wodny, panorama

Opis wygenerowany automatycznie

potok\_3-1 (próbka 10%)

Obraz zawierający statek, statek wodny

Opis wygenerowany automatycznie

potok\_3-2 (próbka 10%)

Obraz zawierający panorama, statek wodny

Opis wygenerowany automatycznie

Osiągane czasy przetwarzania dla wszystkich potoków wyglądały podobnie. Wydaje się, że zarówno w przypadku potok\_1-1 jak i potok\_3-1 zdarzenie odstające występują nieco rzadziej aniżeli w przypadku potok\_1-2, potok\_3-1. Być może istnieje w tym zjawisku interpretacja biznesowo-techniczna, która nie jest mi w tej chwili znana.

Drążenie danych po tygodniu potwierdza, że na przestrzeni obu miesięcy zdarzenia odstające występowały regularnie. Wydaje się, że wstępnie każdy czas procesowania dłuższy aniżeli 10 sekund można uznać za wartość odstającą.

Obraz zawierający tekst, diagram, linia, zrzut ekranu

Opis wygenerowany automatycznie

Obraz zawierający tekst, diagram, zrzut ekranu, linia

Opis wygenerowany automatycznie

Obraz zawierający diagram, linia, tekst, zrzut ekranu

Opis wygenerowany automatycznie

Obraz zawierający diagram, linia, tekst, Wykres

Opis wygenerowany automatycznie

Obraz zawierający tekst, linia, diagram, Wykres

Opis wygenerowany automatycznie

Obraz zawierający diagram, linia, tekst, Wykres

Opis wygenerowany automatycznie

Obraz zawierający diagram, tekst, linia, Wykres

Opis wygenerowany automatycznie

Obraz zawierający diagram, tekst, linia, Wykres

Opis wygenerowany automatycznie

Obraz zawierający tekst, diagram, linia, zrzut ekranu

Opis wygenerowany automatycznie

Z grubsza we wszystkich tygodniach o każdej godzinie doświadczano zdarzeń odstających  
(proszę powiększyć scrollem myszki)

Obraz zawierający tekst, dokument, Czcionka, papier

Opis wygenerowany automatycznie

# Eksploracyjna Analiza Danych – potok\_1-1

## **Zbiór danych** - typy danych

|  |  |
| --- | --- |
| Numeryczna | 6 |
| Kategoria | 2 |
| Data-czas | 1 |

## **Zbiór danych** - generalne statystyki

|  |  |
| --- | --- |
| Liczba zmiennych | 9 |
| Liczba obserwacji | 271952 |
| Braki danych | 32681 |
| Braki danych (%) | 1.3% |
| Zduplikowane wiersze | 0 |
| Zduplikowane wiersze (%) | 0.0% |

## **Zbiór danych** - ostrzeżenia

|  |  |
| --- | --- |
| [ID\_Process](http://localhost:63342/yet-another-analysis/outputs/report_potok_1_1.html?_ijt=k6ns0llggrv4e93d61ij2rga92&_ij_reload=RELOAD_ON_SAVE#pp_var_4115601506032977688) jest silnie skorelowany z ID\_Document | **High correlation** |
| [ID\_Document](http://localhost:63342/yet-another-analysis/outputs/report_potok_1_1.html?_ijt=k6ns0llggrv4e93d61ij2rga92&_ij_reload=RELOAD_ON_SAVE#pp_var_4369173452904466745) jest silnie skorelowany z ID\_Process | **High correlation** |
| [ProcessingDuration\_sec](http://localhost:63342/yet-another-analysis/outputs/report_potok_1_1.html?_ijt=k6ns0llggrv4e93d61ij2rga92&_ij_reload=RELOAD_ON_SAVE#pp_var_5075829074120433651) jest silnie skorelowany z Duration\_sec | **High correlation** |
| [Duration\_sec](http://localhost:63342/yet-another-analysis/outputs/report_potok_1_1.html?_ijt=k6ns0llggrv4e93d61ij2rga92&_ij_reload=RELOAD_ON_SAVE#pp_var_-3689515850995436885) jest silnie skorelowany z ProcessingDuration\_sec | **High correlation** |
| [Retries](http://localhost:63342/yet-another-analysis/outputs/report_potok_1_1.html?_ijt=k6ns0llggrv4e93d61ij2rga92&_ij_reload=RELOAD_ON_SAVE#pp_var_4593143894907911153) jest silnie niezbilansowane (98.6%) | **Imbalance** |
| [InputPageCount](http://localhost:63342/yet-another-analysis/outputs/report_potok_1_1.html?_ijt=k6ns0llggrv4e93d61ij2rga92&_ij_reload=RELOAD_ON_SAVE#pp_var_-3646287023250528572) ma 10562 (3.9%) braków danych | **Missing** |
| [ProcessingDuration\_sec](http://localhost:63342/yet-another-analysis/outputs/report_potok_1_1.html?_ijt=k6ns0llggrv4e93d61ij2rga92&_ij_reload=RELOAD_ON_SAVE#pp_var_5075829074120433651) ma 10562 (3.9%) braków danych | **Missing** |
| [Duration\_sec](http://localhost:63342/yet-another-analysis/outputs/report_potok_1_1.html?_ijt=k6ns0llggrv4e93d61ij2rga92&_ij_reload=RELOAD_ON_SAVE#pp_var_-3689515850995436885) ma 11557 (4.2%) braków danych | **Missing** |
| [InputPageCount](http://localhost:63342/yet-another-analysis/outputs/report_potok_1_1.html?_ijt=k6ns0llggrv4e93d61ij2rga92&_ij_reload=RELOAD_ON_SAVE#pp_var_-3646287023250528572) ma silnie skośny rozkład (γ1 = 206.1385284) | **Skewed** |
| [ID\_Process](http://localhost:63342/yet-another-analysis/outputs/report_potok_1_1.html?_ijt=k6ns0llggrv4e93d61ij2rga92&_ij_reload=RELOAD_ON_SAVE#pp_var_4115601506032977688) ma unikatowe wartości | **Unique** |
| [MessagesInQueue](http://localhost:63342/yet-another-analysis/outputs/report_potok_1_1.html?_ijt=k6ns0llggrv4e93d61ij2rga92&_ij_reload=RELOAD_ON_SAVE#pp_var_6355307072904448556) ma 43941 (16.2%) zer | **Zeros** |

## [**Zmienna - InputPageCount**](http://localhost:63342/yet-another-analysis/outputs/report_timeseries.html?_ijt=tapd920bd1e4gb0nrgmalqbhcu&_ij_reload=RELOAD_ON_SAVE#pp_var_8935948530309897777) - charakterystyki zmiennej

|  |  |
| --- | --- |
| Wartości unikatowe | 52 |
| Wartości unikatowe **(%)** | < 0.1% |
| Braki danych | 10562 |
| Braki danych **(%)** | 3.9% |
| **Średnia** | 1.1110371 |
| **Minimum** | 1 |
| **Maksimum** | 530 |
| **Liczba zer** | 0 |
| **Liczba zer (%)** | 0.0% |
| **Liczba wartości negatywnych** | 0 |
| **Liczba wartości negatywnych (%)** | 0.0% |

Statystyki kwantylowe

|  |  |
| --- | --- |
| **Minimum** | 1 |
| **5-th percentyl** | 1 |
| **Q1** | 1 |
| **mediana** | 1 |
| **Q3** | 1 |
| **95-th percentyl** | 2 |
| **Maksimum** | 530 |
| **Rozstęp** | 529 |
| **Rozstęp miedzykwartylowy (IQR)** | 0 |

Statystyki opisowe

|  |  |
| --- | --- |
| **Odchylenie standardowe** | 1.8626924 |
| **Współczynnik wariancji (CV)** | 1.6765347 |
| **Kurtoza** | 53036.396 |
| **Średnia** | 1.1110371 |
| **Median Absolute Deviation (MAD)** | 0 |
| **Skośność** | 206.13853 |
| **Sum** | 290414 |
| **Wariancja** | 3.4696229 |

Wartości najczęstsze - 10 wartości minimalnych

| **Wartość** | **Liczebność** | **Częstotliwość (%)** |
| --- | --- | --- |
| 1 | 247644 | 91.1% |
| 2 | 9057 | 3.3% |
| 3 | 2329 | 0.9% |
| 4 | 1037 | 0.4% |
| 5 | 353 | 0.1% |
| 6 | 352 | 0.1% |
| 7 | 143 | 0.1% |
| 8 | 118 | < 0.1% |
| 9 | 51 | < 0.1% |
| 10 | 42 | < 0.1% |

Wartości ekstremalne - 10 wartości maksymalnych

|  |
| --- |
|  |
| **Wartość** | **Liczebność** | **Częstotliwość** (%) |
| 530 | 1 | < 0.1% |
| 459 | 2 | < 0.1% |
| 150 | 1 | < 0.1% |
| 86 | 1 | < 0.1% |
| 74 | 2 | < 0.1% |
| 70 | 1 | < 0.1% |
| 62 | 1 | < 0.1% |
| 61 | 2 | < 0.1% |
| 53 | 1 | < 0.1% |
| 51 | 1 | < 0.1% |

## **Zmienna - MessagesInQueue** - charakterystyki zmiennej

|  |  |
| --- | --- |
| Wartości unikatowe | 88 |
| Wartości unikatowe **(%)** | < 0.1% |
| Braki danych | 0 |
| Braki danych **(%)** | 0.0% |
| **Średnia** | 3.7300884 |
| **Minimum** | 0 |
| **Maksimum** | 88 |
| **Liczba zer** | 43941 |
| **Liczba zer (%)** | 16.2% |
| **Liczba wartości negatywnych** | 0 |
| **Liczba wartości negatywnych (%)** | 0.0% |

Statystyki kwantylowe

|  |  |
| --- | --- |
| **Minimum** | 0 |
| **5-th percentyl** | 0 |
| **Q1** | 1 |
| **mediana** | 2 |
| **Q3** | 5 |
| **95-th percentyl** | 12 |
| **Maksimum** | 88 |
| **Rozstęp** | 88 |
| **Rozstęp miedzykwartylowy (IQR)** | 4 |

Statystyki opisowe

|  |  |
| --- | --- |
| **Odchylenie standardowe** | 4.873005 |
| **Kurtoza** | 30.103881 |
| **Średnia** | 3.7300884 |
| **Median Absolute Deviation (MAD)** | 2 |
| **Skośność** | 4.2227733 |
| **Sum** | 1014405 |
| **Wariancja** | 23.746177 |

Wartości najczęstsze - 10 wartości minimalnych

| **Wartość** | **Liczebność** | **Częstotliwość (%)** |
| --- | --- | --- |
| 0 | 43941 | 16.2% |
| 1 | 44766 | 16.5% |
| 2 | 48054 | 17.7% |
| 3 | 38976 | 14.3% |
| 4 | 27374 | 10.1% |
| 5 | 17979 | 6.6% |
| 6 | 12001 | 4.4% |
| 7 | 8335 | 3.1% |
| 8 | 6023 | 2.2% |
| 9 | 4330 | 1.6% |

Wartości ekstremalne - 10 wartości maksymalnych

|  |
| --- |
|  |
| **Wartość** | **Liczebność** | **Częstotliwość** (%) |
| 88 | 2 | < 0.1% |
| 87 | 1 | < 0.1% |
| 86 | 1 | < 0.1% |
| 85 | 1 | < 0.1% |
| 84 | 1 | < 0.1% |
| 82 | 2 | < 0.1% |
| 81 | 1 | < 0.1% |
| 80 | 1 | < 0.1% |
| 79 | 2 | < 0.1% |
| 78 | 2 | < 0.1% |

## **Zmienna - Retries** - charakterystyki zmiennej

|  |  |
| --- | --- |
| Wartości unikatowe | 3 |
| Wartości unikatowe **(%)** | < 0.1% |
| Braki danych | 0 |
| Braki danych **(%)** | 0.0% |

Wartości najczęstsze

| **Wartość** | **Liczebność** | **Częstotliwość (%)** |
| --- | --- | --- |
| 0 | 271371 | 99.8% |
| 1 | 573 | 0.2% |
| 2 | 8 | < 0.1% |

## **Zmienna - ProcessingDuration\_sec** - charakterystyki zmiennej

|  |  |
| --- | --- |
| Wartości unikatowe | 13556 |
| Wartości unikatowe **(%)** | 5.2% |
| Braki danych | 10562 |
| Braki danych **(%)** | 3.9% |
| **Średnia** | 3.1910448 |
| **Minimum** | 0.039 |
| **Maksimum** | 175.24 |
| **Liczba zer** | 0 |
| **Liczba zer (%)** | 0.0% |
| **Liczba wartości negatywnych** | 0 |
| **Liczba wartości negatywnych (%)** | 0.0% |

Statystyki kwantylowe

|  |  |
| --- | --- |
| **Minimum** | 0.039 |
| **5-th percentyl** | 0.108 |
| **Q1** | 2.192 |
| **mediana** | 2.589 |
| **Q3** | 3.506 |
| **95-th percentyl** | 7.126 |
| **Maksimum** | 175.24 |
| **Rozstęp** | 175.24 |
| **Rozstęp miedzykwartylowy (IQR)** | 1.314 |

Statystyki opisowe

|  |  |
| --- | --- |
| **Odchylenie standardowe** | 2.8108951 |
| **Współczynnik wariancji (CV)** | 0.88086982 |
| **Kurtoza** | 273.85347 |
| **Średnia** | 3.1910448 |
| **Median Absolute Deviation (MAD)** | 0.506 |
| **Skośność** | 9.9315516 |
| **Sum** | 7.901131 |
| **Wariancja** | 26.993447 |

Wartości najczęstsze

| **Wartość** | **Liczebność** | **Częstotliwość (%)** |
| --- | --- | --- |
| 0.039 | 1 | <0.1% |
| 0.04 | 1 | <0.1% |
| 0.044 | 2 | <0.1% |
| 0.045 | 2 | <0.1% |
| 0.045 | 2 | <0.1% |
| 0.047 | 5 | <0.1% |
| 0.048 | 9 | <0.1% |
| 0.049 | 6 | <0.1% |
| 0.05 | 10 | <0.1% |

Wartości ekstremalne - 10 wartości maksymalnych

|  |
| --- |
|  |
| **Wartość** | **Liczebność** | **Częstotliwość** (%) |
| 175.24 | 1 | < 0.1% |
| 145.442 | 1 | < 0.1% |
| 130.966 | 1 | < 0.1% |
| 121.274 | 1 | < 0.1% |
| 119.344 | 1 | < 0.1% |
| 119.04 | 1 | < 0.1% |
| 119.04 | 1 | < 0.1% |
| 111.732 | 1 | < 0.1% |
| 107.177 | 1 | < 0.1% |
| 106.438 | 1 | < 0.1% |

## **Zmienna - Duration\_sec** - charakterystyki zmiennej

|  |  |
| --- | --- |
| Wartości unikatowe | 18711 |
| Wartości unikatowe **(%)** | 7.2% |
| Braki danych | 11557 |
| Braki danych **(%)** | 4.2% |
| **Średnia** | 5.6966176 |
| **Minimum** | 0.295 |
| **Maksimum** | 676.629 |
| **Liczba zer** | 0 |
| **Liczba zer (%)** | 0.0% |
| **Liczba wartości negatywnych** | 0 |
| **Liczba wartości negatywnych (%)** | 0.0% |

Statystyki kwantylowe

|  |  |
| --- | --- |
| **Minimum** | 0.295 |
| **5-th percentyl** | 0.537 |
| **Q1** | 2.598 |
| **mediana** | 3.138 |
| **Q3** | 4.543 |
| **95-th percentyl** | 10.792 |
| **Maksimum** | 676.629 |
| **Rozstęp** | 676.629 |
| **Rozstęp miedzykwartylowy (IQR)** | 1.945 |

Statystyki opisowe

|  |  |
| --- | --- |
| **Odchylenie standardowe** | 22.285425 |
| **Współczynnik wariancji (CV)** | 3.9120451 |
| **Kurtoza** | 207.67929 |
| **Średnia** | 207.67929 |
| **Median Absolute Deviation (MAD)** | 0.704 |
| **Skośność** | 14.094964 |
| **Sum** | 1483370.7 |
| **Wariancja** | 496.64016 |

Wartości najczęstsze

| **Wartość** | **Liczebność** | **Częstotliwość (%)** |
| --- | --- | --- |
| 0.295 | 2 | <0.1% |
| 0.297 | 1 | <0.1% |
| 0.3 | 1 | <0.1% |
| 0.301 | 1 | <0.1% |
| 0.302 | 1 | <0.1% |
| 0.306 | 2 | <0.1% |
| 0.307 | 3 | <0.1% |
| 0.309 | 2 | <0.1% |
| 0.311 | 4 | <0.1% |

Wartości ekstremalne - 10 wartości maksymalnych

|  |
| --- |
|  |
| **Wartość** | **Liczebność** | **Częstotliwość** (%) |
| 676.629 | 1 | < 0.1% |
| 631.244 | 1 | < 0.1% |
| 623.835 | 1 | < 0.1% |
| 617.84 | 1 | < 0.1% |
| 617.284 | 4 | < 0.1% |
| 606.682 | 1 | < 0.1% |
| 381.898 | 3 | < 0.1% |
| 380.301 | 1 | < 0.1% |
| 369.181 | 1 | < 0.1% |
| 358.973 | 1 | < 0.1% |

## **Zmienna - StartDateTimeUtc** - charakterystyki zmiennej

|  |  |
| --- | --- |
| Wartości unikatowe | 271000 |
| Wartości unikatowe **(%)** | 99.6% |
| Braki danych | 0 |
| Braki danych **(%)** | 0.0% |
| **Minimum** | 2023-02-01 06:40:34.927000 |
| **Maksimum** | 2023-03-13 09:33:49.717000 |

Obraz zawierający tekst, zrzut ekranu, Czcionka, diagram

Opis wygenerowany automatycznie

## **Tablice korelacji** – korelacja między zmiennymi

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **InputPageCount** | **MessagesInQueue** | **ProcessingDuration\_sec** | **Duration\_sec** | **Retries** |  |
| **InputPageCount** | 1 | 0,002 | 0,223 | 0,201 | 0 |
| **MessagesInQueue** | 0,002 | 1 | 0,34 | 0,358 | 0,053 |
| **ProcessingDuration\_sec** | 0,223 | 0,34 | 1 | 0,894 | 0,008 |
| **Duration\_sec** | 0,201 | 0,358 | 0,894 | 1 | 0,198 |
| **Retries** | 0 | 0,053 | 0,008 | 0,198 | 1 |

Obraz zawierający tekst, zrzut ekranu, diagram, numer

Opis wygenerowany automatycznie

## **Tablice korelacji** – korelacja między brakami danych pomiędzy zmiennymi

Obraz zawierający tekst, zrzut ekranu, diagram

Opis wygenerowany automatycznie

# Eksploracyjna Analiza Danych – potok\_1-2

# Eksploracyjna Analiza Danych – potok\_3-1

# Eksploracyjna Analiza Danych – potok\_3-2

## **Zbiór danych** - typy danych

|  |  |
| --- | --- |
| Numeryczna | 6 |
| Kategoria | 2 |
| Data-czas | 1 |

## **Zbiór danych** - generalne statystyki

|  |  |
| --- | --- |
| Liczba zmiennych | 9 |
| Liczba obserwacji | 245732 |
| Braki danych | 19746 |
| Braki danych (%) | 1.1% |
| Zduplikowane wiersze | 11 |
| Zduplikowane wiersze (%) | < 0.1% |

## **Zbiór danych** - ostrzeżenia

|  |  |
| --- | --- |
| Zbiór danych ma 11 (< 0.1%) duplikatów | **High correlation** |
| [MessagesInQueue](http://localhost:63342/yet-another-analysis/outputs/report_potok_3_2.html?_ijt=n5c28ijbh6pm3efgniq0h5jcsq&_ij_reload=RELOAD_ON_SAVE#pp_var_621005032818396674) jest silnie skorelowana z Duration\_sec | **High correlation** |
| [ProcessingDuration\_sec](http://localhost:63342/yet-another-analysis/outputs/report_potok_3_2.html?_ijt=n5c28ijbh6pm3efgniq0h5jcsq&_ij_reload=RELOAD_ON_SAVE#pp_var_-3653951573983100350) jest silnie skorelowana z Duration\_sec | **High correlation** |
| [Duration\_sec](http://localhost:63342/yet-another-analysis/outputs/report_potok_3_2.html?_ijt=n5c28ijbh6pm3efgniq0h5jcsq&_ij_reload=RELOAD_ON_SAVE#pp_var_-8521686961172518299) jest silnie skorelowana z MessagesInQueue and 1 other fields | **Imbalance** |
| [InputPageCount](http://localhost:63342/yet-another-analysis/outputs/report_potok_3_2.html?_ijt=n5c28ijbh6pm3efgniq0h5jcsq&_ij_reload=RELOAD_ON_SAVE#pp_var_-1001698636911930012) ma 5966 (2.4%) braków danych | **Missing** |
| [ProcessingDuration\_sec](http://localhost:63342/yet-another-analysis/outputs/report_potok_3_2.html?_ijt=n5c28ijbh6pm3efgniq0h5jcsq&_ij_reload=RELOAD_ON_SAVE#pp_var_-3653951573983100350) ma 5966 (2.4%) braków danych | **Missing** |
| [Duration\_sec](http://localhost:63342/yet-another-analysis/outputs/report_potok_3_2.html?_ijt=n5c28ijbh6pm3efgniq0h5jcsq&_ij_reload=RELOAD_ON_SAVE#pp_var_-8521686961172518299) ma 7814 (3.2%) braków danych | **Missing** |
| [InputPageCount](http://localhost:63342/yet-another-analysis/outputs/report_potok_3_2.html?_ijt=n5c28ijbh6pm3efgniq0h5jcsq&_ij_reload=RELOAD_ON_SAVE#pp_var_-1001698636911930012) charakteryzuje silnie skośny rozkład (γ1 = 89.26361956) | **Skewed** |
| [Retries](http://localhost:63342/yet-another-analysis/outputs/report_potok_3_2.html?_ijt=n5c28ijbh6pm3efgniq0h5jcsq&_ij_reload=RELOAD_ON_SAVE#pp_var_-8561503850158418306) charakteryzuje silnie skośny rozkład (γ1 = 21.83594329) | **Unique** |
| [ProcessingDuration\_sec](http://localhost:63342/yet-another-analysis/outputs/report_potok_3_2.html?_ijt=n5c28ijbh6pm3efgniq0h5jcsq&_ij_reload=RELOAD_ON_SAVE#pp_var_-3653951573983100350) charakteryzuje silnie skośny rozkład (γ1 = 51.66540777) | **Zeros** |

## [**Zmienna - InputPageCount**](http://localhost:63342/yet-another-analysis/outputs/report_timeseries.html?_ijt=tapd920bd1e4gb0nrgmalqbhcu&_ij_reload=RELOAD_ON_SAVE#pp_var_8935948530309897777) - charakterystyki zmiennej

|  |  |
| --- | --- |
| Wartości unikatowe | 74 |
| Wartości unikatowe **(%)** | < 0.1% |
| Braki danych | 5966 |
| Braki danych **(%)** | 2.4% |
| **Średnia** | 1.3099689 |
| **Minimum** | 1 |
| **Maksimum** | 394 |
| **Liczba zer** | 0 |
| **Liczba zer (%)** | 0.0% |
| **Liczba wartości negatywnych** | 0 |
| **Liczba wartości negatywnych (%)** | 0.0% |

Statystyki kwantylowe

|  |  |
| --- | --- |
| **Minimum** | 1 |
| **5-th percentyl** | 1 |
| **Q1** | 1 |
| **mediana** | 1 |
| **Q3** | 1 |
| **95-th percentyl** | 3 |
| **Maksimum** | 394 |
| **Rozstęp** | 393 |
| **Rozstęp miedzykwartylowy (IQR)** | 0 |

Statystyki opisowe

|  |  |
| --- | --- |
| **Odchylenie standardowe** | 2.3113962 |
| **Współczynnik wariancji (CV)** | 1.764466 |
| **Kurtoza** | 12393.494 |
| **Średnia** | 1.3099689 |
| **Median Absolute Deviation (MAD)** | 0 |
| **Skośność** | 89.26362 |
| **Sum** | 314086 |
| **Wariancja** | 5.3425526 |

Wartości najczęstsze - 10 wartości minimalnych

| **Wartość** | **Liczebność** | **Częstotliwość (%)** |
| --- | --- | --- |
| 1 | 206797 | 84.2% |
| 2 | 20067 | 8.2% |
| 3 | 6050 | 2.5% |
| 4 | 2831 | 1.2% |
| 5 | 1033 | 0.4% |
| 6 | 1094 | 0.4% |
| 7 | 411 | 0.2% |
| 8 | 371 | 0.2% |
| 9 | 204 | <0.1% |
| 10 | 171 | 0.1% |

Wartości ekstremalne - 10 wartości maksymalnych

|  |
| --- |
|  |
| **Wartość** | **Liczebność** | **Częstotliwość** (%) |
| 394 | 1 | < 0.1% |
| 355 | 2 | < 0.1% |
| 350 | 1 | < 0.1% |
| 291 | 1 | < 0.1% |
| 242 | 2 | < 0.1% |
| 220 | 1 | < 0.1% |
| 186 | 1 | < 0.1% |
| 106 | 2 | < 0.1% |
| 101 | 1 | < 0.1% |
| 100 | 2 | < 0.1% |

## **Zmienna - MessagesInQueue** - charakterystyki zmiennej

|  |  |
| --- | --- |
| Wartości unikatowe | 1670 |
| Wartości unikatowe **(%)** | 0.7% |
| Braki danych | 0 |
| Braki danych **(%)** | 0.0% |
| **Średnia** | 19.935373 |
| **Minimum** | 2018 |
| **Maksimum** | 88 |
| **Liczba zer** | 66685 |
| **Liczba zer (%)** | 27.1% |
| **Liczba wartości negatywnych** | 0 |
| **Liczba wartości negatywnych (%)** | 0.0% |

Statystyki kwantylowe

|  |  |
| --- | --- |
| **Minimum** | 0 |
| **5-th percentyl** | 0 |
| **Q1** | 0 |
| **mediana** | 3 |
| **Q3** | 12 |
| **95-th percentyl** | 55 |
| **Maksimum** | 2018 |
| **Rozstęp** | 2018 |
| **Rozstęp miedzykwartylowy (IQR)** | 12 |

Statystyki opisowe

|  |  |
| --- | --- |
| **Odchylenie standardowe** | 109.1699 |
| **Kurtoza** | 185.48726 |
| **Średnia** | 19.935373 |
| **Median Absolute Deviation (MAD)** | 2 |
| **Skośność** | 12.935543 |
| **Sum** | 4898759 |
| **Wariancja** | 11918.066 |

Wartości najczęstsze - 10 wartości minimalnych

| **Wartość** | **Liczebność** | **Częstotliwość (%)** |
| --- | --- | --- |
| 0 | 66685 | 27.1% |
| 1 | 30862 | 12.6% |
| 2 | 19473 | 7.9% |
| 3 | 13768 | 5.6% |
| 4 | 11615 | 4.7% |
| 5 | 9316 | 3.8% |
| 6 | 7812 | 3.2% |
| 7 | 6308 | 2.6% |
| 8 | 5647 | 2.3% |
| 9 | 4879 | 2% |

Wartości ekstremalne - 10 wartości maksymalnych

|  |
| --- |
|  |
| **Wartość** | **Liczebność** | **Częstotliwość** (%) |
| 2018 | 1 | < 0.1% |
| 2017 | 1 | < 0.1% |
| 2016 | 1 | < 0.1% |
| 2015 | 1 | < 0.1% |
| 2011 | 4 | < 0.1% |
| 2010 | 1 | < 0.1% |
| 2007 | 3 | < 0.1% |
| 2006 | 1 | < 0.1% |
| 2005 | 1 | < 0.1% |
| 2003 | 1 | < 0.1% |

## **Zmienna - Retries** - charakterystyki zmiennej

|  |  |
| --- | --- |
| Wartości unikatowe | 10 |
| Wartości unikatowe **(%)** | < 0.1% |
| Braki danych | 0 |
| Braki danych **(%)** | 0.0% |

Wartości najczęstsze

Wartości najczęstsze - 10 wartości minimalnych

|  |  |  |
| --- | --- | --- |
| Value | Count | Frequency (%) |
| 0 | 243584 | 99.1% |
| 1 | 1663 |  |
| 0.7% |
| 2 | 374 |  |
| 0.2% |
| 3 | 84 |  |
| < 0.1% |
| 4 | 3 |  |
| < 0.1% |
| 5 | 9 |  |
| < 0.1% |
| 6 | 2 |  |
| < 0.1% |
| 7 | 4 |  |
| < 0.1% |
| 8 | 1 |  |
| < 0.1% |
| 9 | 8 |  |
| < 0.1% |

Wartości ekstremalne - 10 wartości maksymalnych

|  |  |  |
| --- | --- | --- |
| Value | Count | Frequency (%) |
| 9 | 8 |  |
| < 0.1% |
| 8 | 1 |  |
| < 0.1% |
| 7 | 4 |  |
| < 0.1% |
| 6 | 2 |  |
| < 0.1% |
| 5 | 9 |  |
| < 0.1% |
| 4 | 3 |  |
| < 0.1% |
| 3 | 84 |  |
| < 0.1% |
| 2 | 374 |  |
| 0.2% |
| 1 | 1663 |  |
| 0.7% |
| 0 | 243584 | 99.1% |

## **Zmienna - ProcessingDuration\_sec** - charakterystyki zmiennej

|  |  |
| --- | --- |
| Wartości unikatowe | 21541 |
| Wartości unikatowe **(%)** | 9.0% |
| Braki danych | 5966 |
| Braki danych **(%)** | 2.4% |
| **Średnia** | 5.3244567 |
| **Minimum** | 0.034 |
| **Maksimum** | 1422.42 |
| **Liczba zer** | 0 |
| **Liczba zer (%)** | 0.0% |
| **Liczba wartości negatywnych** | 0 |
| **Liczba wartości negatywnych (%)** | 0.0% |

Statystyki kwantylowe

|  |  |
| --- | --- |
| **Minimum** | 0.034 |
| **5-th percentyl** | 0.127 |
| **Q1** | 2.548 |
| **mediana** | 3.822 |
| **Q3** | 6.581 |
| **95-th percentyl** | 14.32475 |
| **Maksimum** | 1422.42 |
| **Rozstęp** | 1422.386 |
| **Rozstęp miedzykwartylowy (IQR)** | 4.033 |

Statystyki opisowe

|  |  |
| --- | --- |
| **Odchylenie standardowe** | 6.5589866 |
| **Współczynnik wariancji (CV)** | 1.2318603 |
| **Kurtoza** | 9552.0622 |
| **Średnia** | 5.3244567 |
| **Median Absolute Deviation (MAD)** | 1.618 |
| **Skośność** | 51.665408 |
| **Sum** | 1276623.7 |
| **Wariancja** | 43.020305 |

Wartości najczęstsze

| **Wartość** | **Liczebność** | **Częstotliwość (%)** |
| --- | --- | --- |
| 0.034 | 1 | <0.1% |
| 0.035 | 1 | <0.1% |
| 0.036 | 1 | <0.1% |
| 0.037 | 4 | <0.1% |
| 0.038 | 5 | <0.1% |
| 0.039 | 4 | <0.1% |
| 0.04 | 2 | <0.1% |
| 0.041 | 5 | <0.1% |
| 0.042 | 8 | <0.1% |

Wartości ekstremalne - 10 wartości maksymalnych

|  |
| --- |
|  |
| **Wartość** | **Liczebność** | **Częstotliwość** (%) |
| 1422.42 | 1 | < 0.1% |
| 620.435 | 1 | < 0.1% |
| 289.376 | 1 | < 0.1% |
| 277.92 | 1 | < 0.1% |
| 267.352 | 1 | < 0.1% |
| 266.048 | 1 | < 0.1% |
| 262.572 | 1 | < 0.1% |
| 262.306 | 1 | < 0.1% |
| 246.455 | 1 | < 0.1% |
| 222.539 | 1 | < 0.1% |

## **Zmienna - Duration\_sec** - charakterystyki zmiennej

|  |  |
| --- | --- |
| Wartości unikatowe | 50609 |
| Wartości unikatowe **(%)** | 21.3% |
| Braki danych | 7814 |
| Braki danych **(%)** | 3.2% |
| **Średnia** | 79.35706 |
| **Minimum** | 0.286 |
| **Maksimum** | 12114.593 |
| **Liczba zer** | 0 |
| **Liczba zer (%)** | 0.0% |
| **Liczba wartości negatywnych** | 0 |
| **Liczba wartości negatywnych (%)** | 0.0% |

Statystyki kwantylowe

|  |  |
| --- | --- |
| **Minimum** | 0.286 |
| **5-th percentyl** | 0.615 |
| **Q1** | 3.105 |
| **mediana** | 5.161 |
| **Q3** | 12.644 |
| **95-th percentyl** | 65.7343 |
| **Maksimum** | 12114.593 |
| **Rozstęp** | 12114.307 |
| **Rozstęp miedzykwartylowy (IQR)** | 9.539 |

Statystyki opisowe

|  |  |
| --- | --- |
| **Odchylenie standardowe** | 625.41879 |
| **Współczynnik wariancji (CV)** | 7.881073 |
| **Kurtoza** | 147.48136 |
| **Średnia** | 79.35706 |
| **Median Absolute Deviation (MAD)** | 2.713 |
| **Skośność** | 11.737471 |
| **Sum** | 18880473 |
| **Wariancja** | 391148.66 |

Wartości najczęstsze

| **Wartość** | **Liczebność** | **Częstotliwość (%)** |
| --- | --- | --- |
| 0.286 | 1 | <0.1% |
| 0.296 | 1 | <0.1% |
| 0.305 | 1 | <0.1% |
| 0.307 | 1 | <0.1% |
| 0.309 | 1 | <0.1% |
| 0.311 | 1 | <0.1% |
| 0.314 | 2 | <0.1% |
| 0.315 | 2 | <0.1% |
| 0.316 | 2 | <0.1% |

Wartości ekstremalne - 10 wartości maksymalnych

|  |
| --- |
|  |
| **Wartość** | **Liczebność** | **Częstotliwość** (%) |
| 12114.593 | 1 | < 0.1% |
| 11740.29 | 1 | < 0.1% |
| 11578.316 | 1 | < 0.1% |
| 11383.363 | 1 | < 0.1% |
| 11382.838 | 4 | < 0.1% |
| 11301.524 | 1 | < 0.1% |
| 11253.692 | 3 | < 0.1% |
| 11190.001 | 1 | < 0.1% |
| 11188.385 | 1 | < 0.1% |
| 11187.848 | 1 | < 0.1% |

## **Zmienna - StartDateTimeUtc** - charakterystyki zmiennej

|  |  |
| --- | --- |
| Wartości unikatowe | 242789 |
| Wartości unikatowe **(%)** | 98.8% |
| Braki danych | 0 |
| Braki danych **(%)** | 0.0% |
| **Minimum** | 2023-02-01 00:24:13.247000 |
| **Maksimum** | 2023-03-31 23:53:31.460000 |

Obraz zawierający tekst, zrzut ekranu, Czcionka, diagram

Opis wygenerowany automatycznie

## **Tablice korelacji** – korelacja między zmiennymi

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **InputPageCount** | **MessagesInQueue** | **Retries** | **ProcessingDuration\_sec** | **Duration\_sec** |  |
| **InputPageCount** | 1 | -0,004 | 0,02 | 0,278 | 0,187 |
| **MessagesInQueue** | -0,004 | 1 | 0,11 | 0,443 | 0,566 |
| **Retries** | 0,02 | 0,11 | 1 | 0,086 | 0,156 |
| **ProcessingDuration\_sec** | 0,278 | 0,443 | 0,086 | 1 | 0,832 |
| **Duration\_sec** | 0,187 | 0,566 | 0,156 | 0,832 | 1 |

Obraz zawierający tekst, zrzut ekranu, diagram, kwadrat

Opis wygenerowany automatycznie

## **Tablice korelacji** – korelacja między brakami danych pomiędzy zmiennymi

Obraz zawierający tekst, zrzut ekranu, diagram, linia

Opis wygenerowany automatycznie