ALATI ZA OBRADU PODATAKA

Vizuelizacija kategorijskih podataka

Tabele frekvencija, stubičasti dijagram, kružni dijagram i Pareto dijagram

Predavač: dr Dušan Stefanović

OD PODATAKA

Audi 40
BMW 35
Mercedes 25

DO GRAFIKONA

poređenje i udeo u celini

Šta treba razumeti

ISHODI UČENJA

1. kako nastaje tabela frekvencija
2. razliku između frekvencije i relativne frekvencije
3. kada koristimo stubičasti, a kada kružni dijagram
4. kako se računa kumulativni procenat
5. šta Pareto dijagram dodaje običnom stubičastom dijagramu
6. kako izabrati prikaz za konkretan problem

Vizuelni prikaz ubrzava poređenje kategorija

ZAŠTO VIZUELIZUJEMO PODATKE

Sirovi zapisi

Audi, BMW, Audi, Mercedes, BMW, Audi, BMW, Audi, Mercedes, Audi ...

Sažet prikaz

Audi 40
BMW 35
Mercedes 25

Zaključak

Audi ima najveću prodaju i čini 40% ukupnog broja prodatih vozila.

Vizuelizacija ne menja podatke. Ona olakšava uočavanje odnosa.

Od zapisa do odluke

TOK PRIPREME PRIKAZA

1

SIROVI PODACI

svaki prodati automobil je jedan zapis

2

GRUPISANJE

zapise grupišemo prema marki

3

FREKVENCIJA

brojimo zapise u svakoj grupi

4

PRIKAZ

biramo tabelu ili odgovarajući grafikon

Svaki korak mora biti proverljiv: zbir frekvencija mora odgovarati broju zapisa.

Primer: prodaja 100 automobila

POČETNI SKUP PODATAKA

Prodavac je tokom posmatranog perioda prodao:

AUDI	BMW	MERCEDES
40 vozila	35 vozila	25 vozila

Ukupan broj vozila: 40 + 35 + 25 = 100

Tabela frekvencija sažima broj pojavljivanja

TABELA FREKVENCIJA

Marka	Frekvencija
Audi	40
BMW	35
Mercedes	25
Ukupno	100

Frekvencija je broj zapisa koji pripadaju određenoj kategoriji.

Frekvencija i relativna frekvencija odgovaraju na različita pitanja

OSNOVNI POJMOVI

FREKVENCIJA

Koliko puta se kategorija pojavila?

Audi: 40 vozila

RELATIVNA FREKVENCIJA

Koliki deo ukupnog broja pripada kategoriji?

Audi: 40%

Frekvencija koristi broj zapisa. Relativna frekvencija koristi udeo ili procenat.

Relativna frekvencija povezuje kategoriju sa celinom

PRORAČUN RELATIVNE FREKVENCIJE

$$\text{Relativna frekvencija kategorije} = \frac{\text{frekvencija kategorije}}{\text{ukupan broj zapisa}}$$

Audi: 40 / 100 = 0,40 = 40%

Zbir relativnih frekvencija svih kategorija mora biti 1, odnosno 100%.

Proširena tabela prikazuje broj i udeo

FREKVENCIJA I PROCENAT

Marka	Frekvencija	Relativna frekvencija	Procenat
Audi	40	0,40	40%
BMW	35	0,35	35%
Mercedes	25	0,25	25%
Ukupno	100	1,00	100%

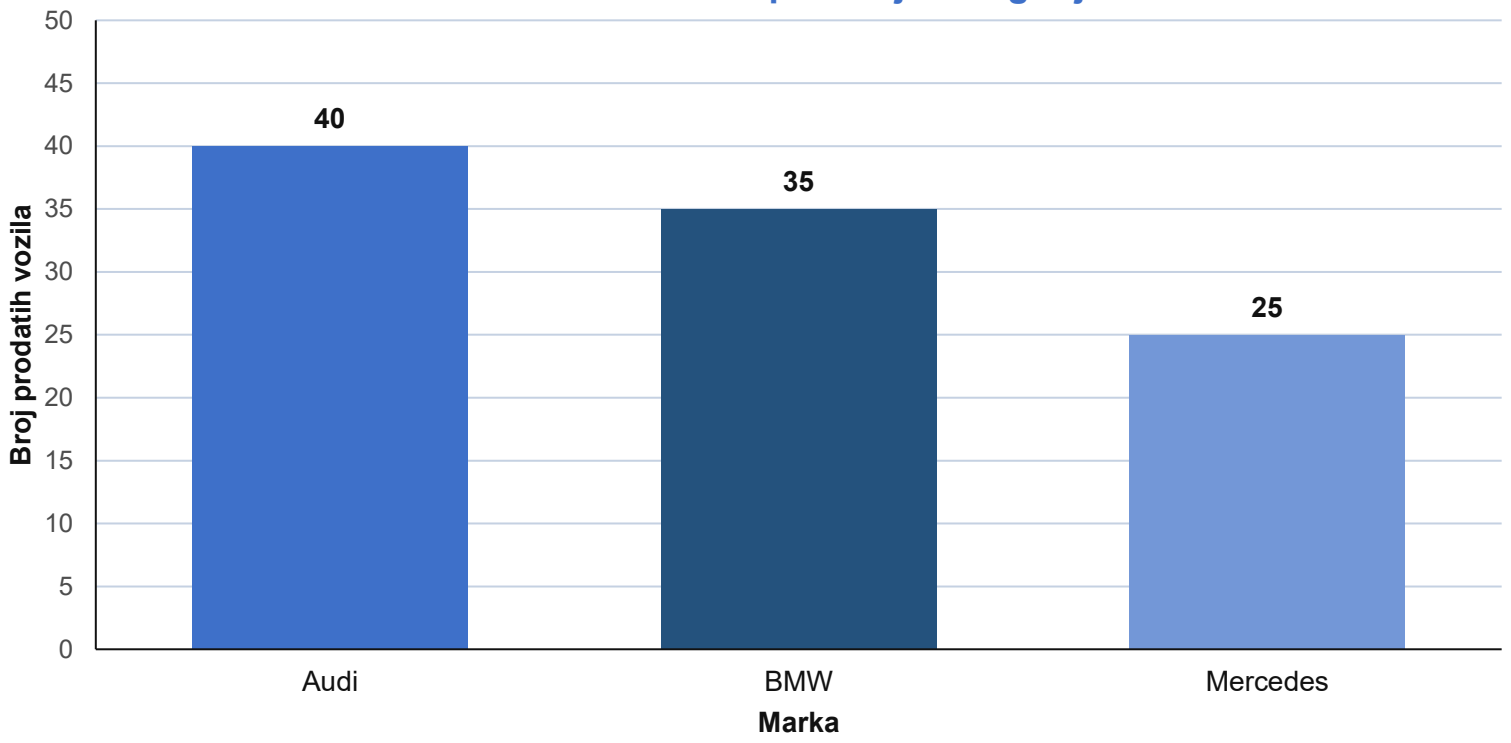
Tri kontrole: 40 + 35 + 25 = 100, 0,40 + 0,35 + 0,25 = 1,00 i 40% + 35% + 25% = 100%.

Stubičasti dijagram poredi frekvencije kategorija

ELEMENTI GRAFIKONA

Horizontalna osa prikazuje kategorije

Visina stuba prikazuje frekvenciju



Razmak između stubova naglašava da su kategorije odvojene.

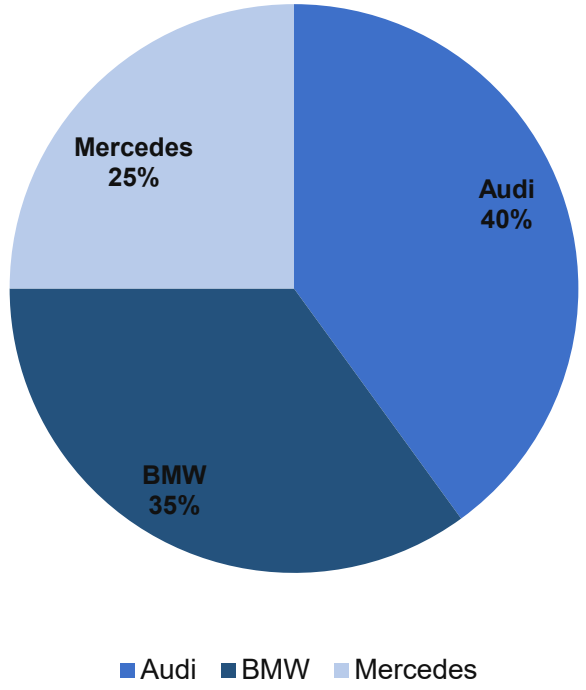
Kružni dijagram zahteva relativne frekvencije

PRIPREMA KRUŽNOG DIJAGRAMA

40 / 100 = 40%

35 / 100 = 35%

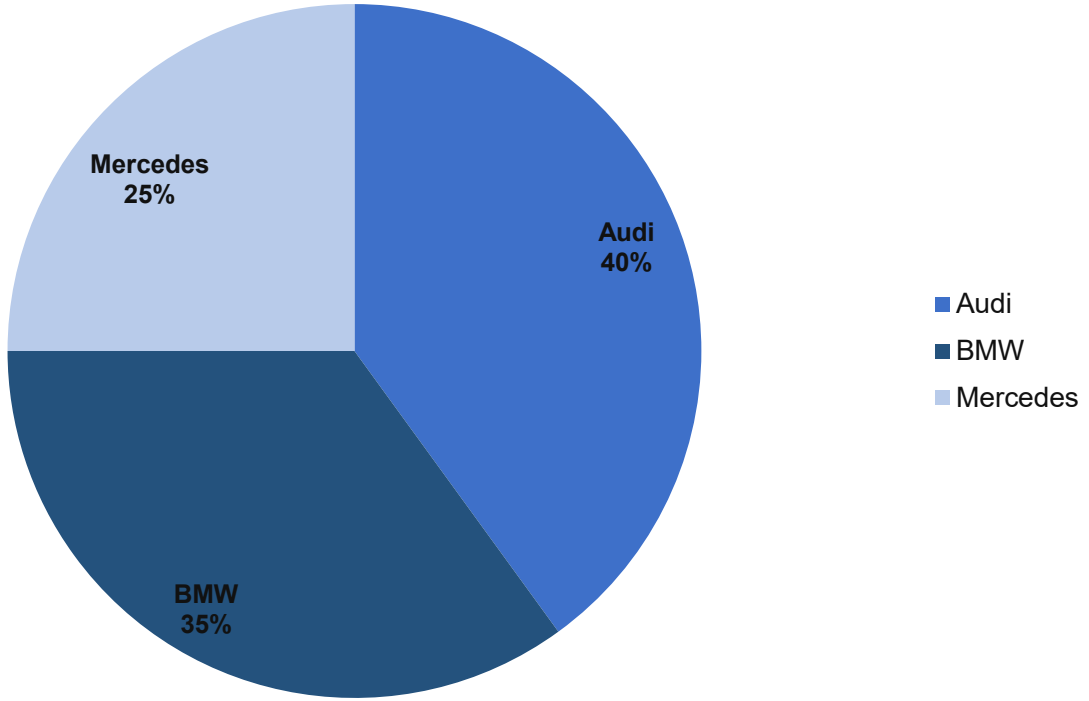
25 / 100 = 25%



Svaki isečak predstavlja udeo kategorije u ukupnom broju zapisa.

Audi čini 40% ukupne prodaje

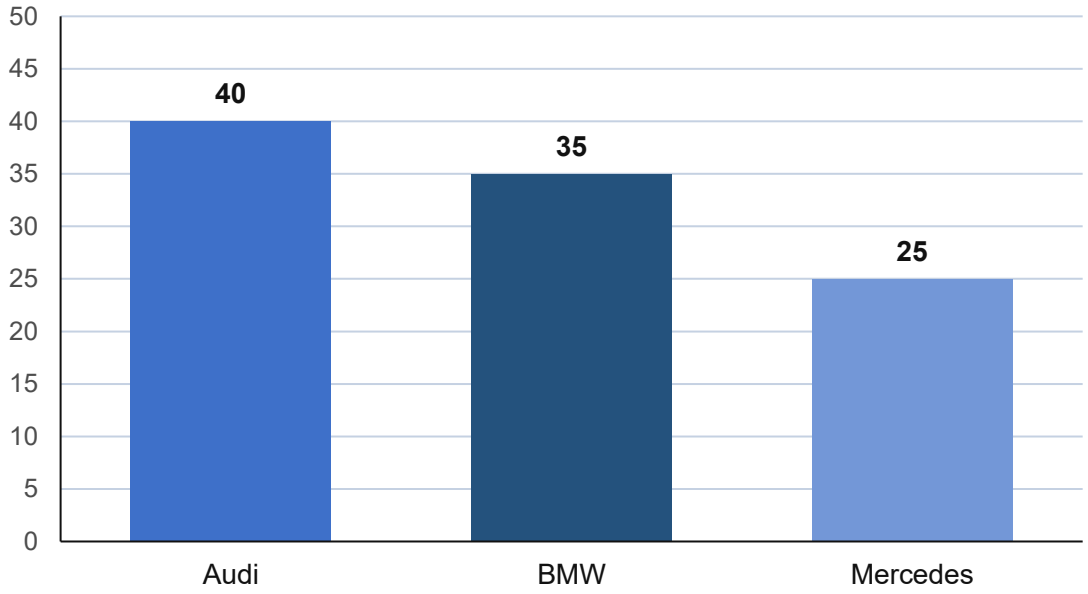
TUMAČENJE KRUŽNOG DIJAGRAMA

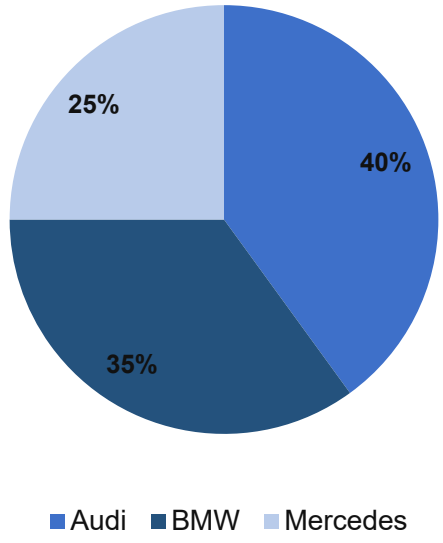


Kružni dijagram naglašava odnos dela i celine, a ne preciznu razliku između sličnih vrednosti.

Stubičasti i kružni dijagram naglašavaju različite odnose

POREĐENJE PRIKAZA





Lakše poređenje tačnih razlika

Lakše sagledavanje udela u celini

Isti podaci mogu dati dva ispravna grafikona, ali grafikoni ne odgovaraju na isto pitanje.

Pareto dijagram kombinuje frekvenciju i kumulativni procenat

PARETO DIJAGRAM

STUBOVI

Kategorije su poređane od najveće ka najmanjoj frekvenciji.

LINIJA

Prikazuje kumulativni procenat nakon dodavanja svake kategorije.

DVE OSE

Leva osa meri frekvenciju. Desna osa meri procenat od 0% do 100%.

Pareto dijagram pokazuje koje kategorije najviše doprinose ukupnom rezultatu.

Pareto analiza počinje sortiranjem frekvencija

PRVI KORAK

Marka	Frekvencija	Redosled
Audi	40	1
BMW	35	2
Mercedes	25	3

Kategorija sa najvećom frekvencijom nalazi se prva, a najmanja poslednja.

Kumulativni procenat raste do 100%

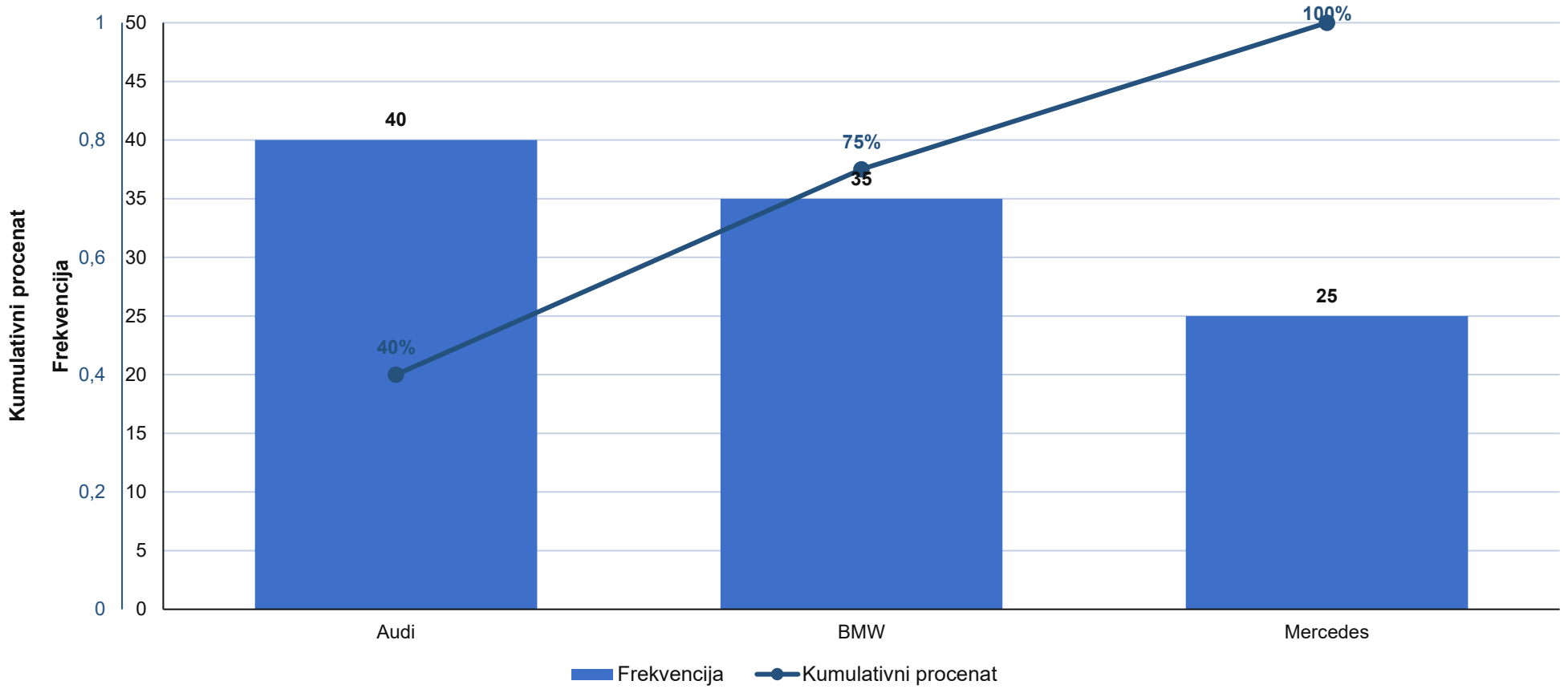
DRUGI KORAK

Marka	Frekvencija	Relativna frekvencija	Kumulativni procenat
Audi	40	40%	40%
BMW	35	35%	40% + 35% = 75%
Mercedes	25	25%	75% + 25% = 100%

Svaka nova tačka linije uključuje trenutnu kategoriju i sve kategorije levo od nje.

Prve dve marke čine 75% posmatrane prodaje

TUMAČENJE PARETO DIJAGRAMA



Linija omogućava da odmah očitamo zajednički udeo prvih nekoliko kategorija.

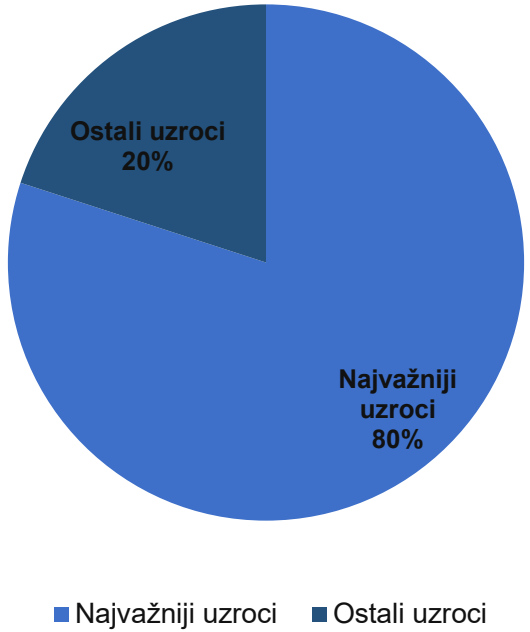
Pareto princip je praktično pravilo, a ne strogi zakon

PRAVILO 80/20

Ilustrativni primer

Od 10 vrsta problema, dve vrste mogu izazvati 80 od 100 prijavljenih incidenata.

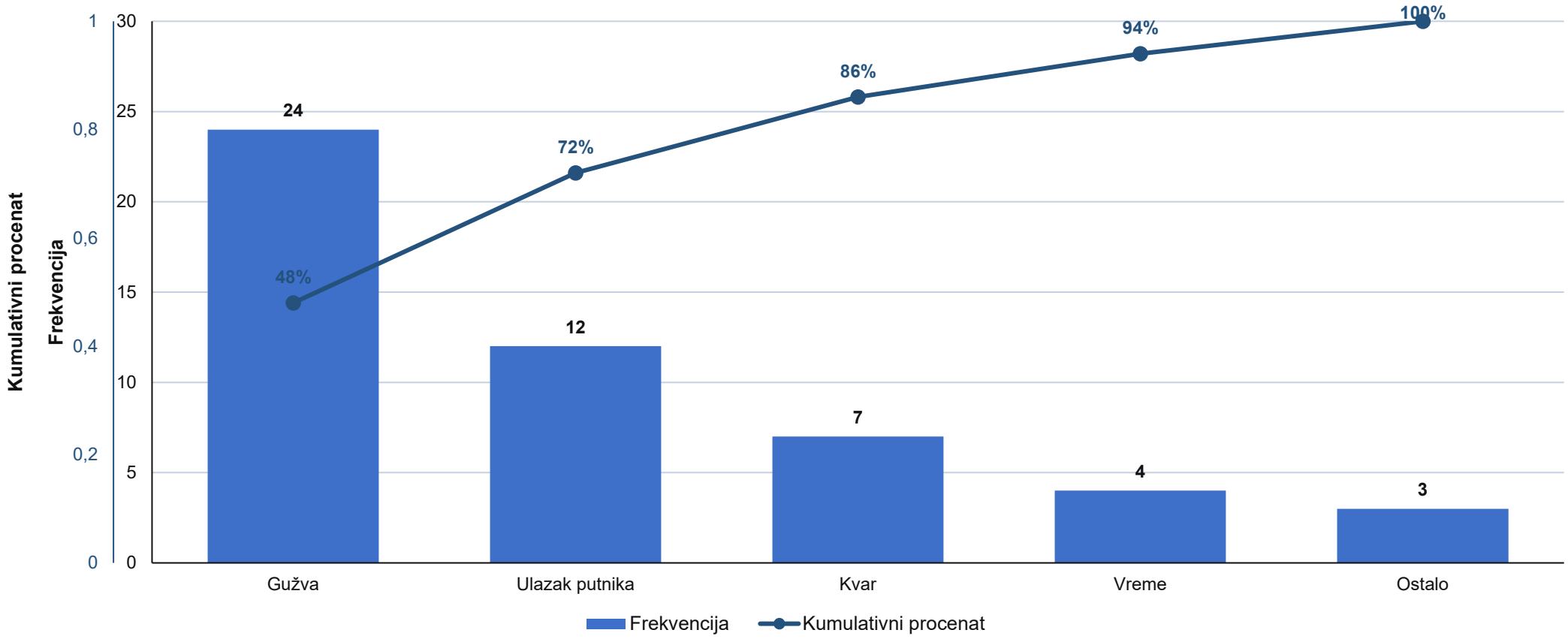
Prioritet dobijaju uzroci koji stvaraju najveći deo posledica.



Odnos u realnim podacima ne mora biti tačno 80% prema 20%.

Primer iz saobraćaja: uzroci kašnjenja autobusa

PRIMENA PARETO DIJAGRAMA



Prva tri uzroka objašnjavaju 86% od 50 zabeleženih kašnjenja.

Izbor prikaza zavisi od pitanja

SAŽETAK I ČESTE GREŠKE

Pitanje	Prikaz
Koliko zapisa pripada kategoriji?	**Tabela frekvencija**
Koja kategorija je veća?	**Stubičasti dijagram**
Koliki je udeo u celini?	**Kružni dijagram**
Koje kategorije daju najveći zajednički doprinos?	**Pareto dijagram**

ČESTE GREŠKE

previše kategorija u kružnom dijagramu

procenti se ne sabiraju do 100%

Pareto stubovi nisu sortirani

presečena osa preuveličava razlike

Prvo definišemo pitanje, zatim biramo prikaz koji najjasnije daje odgovor.

Pitanje 1

PROVERA ZNANJA

Šta predstavlja frekvencija kategorije Audi?

A prosečnu cenu Audija

B broj prodatih Audi vozila

C udeo Audija bez ukupnog broja

D redosled marke u tabeli

Izaberite jedan odgovor i obrazložite izbor.

Pitanje 2

PROVERA ZNANJA

Ako je 35 od 100 prodatih vozila marke BMW, relativna frekvencija iznosi:

A 0,035, odnosno 3,5%

B 0,35, odnosno 35%

C 35 prodatih vozila

D 65% kao ostatak

Izaberite jedan odgovor i obrazložite izbor.

Pitanje 3

PROVERA ZNANJA

Koji grafikon najjasnije pokazuje da je razlika između Audija i BMW-a pet vozila?

A stubičasti dijagram

B kružni dijagram

C samo tekstualna lista

D nijedan prikaz

Izaberite jedan odgovor i obrazložite izbor.

Pitanje 4

PROVERA ZNANJA

Kumulativni procenat posle kategorija od 40% i 35% iznosi:

A Samo prva kategorija: 40%

B Samo druga kategorija: 35%

C Zbir obe kategorije: 75%

D Sve kategorije: 100%

Izaberite jedan odgovor i obrazložite izbor.

Pitanje 5

PROVERA ZNANJA

Koji prikaz je najpogodniji za izdvajanje uzroka koji stvaraju najveći deo kašnjenja?

A Pareto dijagram

B kružni dijagram sa 20 isečaka

C nasumično poređana tabela

D linijski grafikon kroz vreme

Izaberite jedan odgovor i obrazložite izbor.