### Ime, priimek:

LV9: Uporabniški poskus in govorne storitve

Na primeru uporabe glasovnih storitev Amazon Aleksa izvedemo poskus z uporabniki, in analiziramo rezultate.

## Uporabniški poskus

Kot uporabniki bomo preskusili govorno storitev dveh sistemov, Amazon Aleksa, in Google Echo.

Primerjali bomo uporabnost obeh sistemov (usability), ter kvaliteto pogovorne storitve.

Izberite si svojo temo pogovora. Od pogovornega sistema želite izvedeti neke uporabne informacije.

Primer:

potujem turistično v London (izberite mesto, državo : USA, VB..).

ogledati si želim galerijo (znamenitosti, nogometno tekmo, koncert ... )

Rezervirati želim hotel, zvedeti želim vremensko napoved, ...

Tematika, cilj vašega pogovora :

## Izvedba uporabniškega poskusa

Izvedite kratek pogovor na vašo tematiko s pogovornim sistemom, pri tem pa spremljajte odziv sistema, skušajte izvedeti iskane informacije.

* pogovor s sistemom A (Aleksa)
* pogovor s sistemom B (Google Echo)

Pri tem bodite tudi pozorni na kvaliteto pogovora (zgovornost sistema, prijaznost, glas .. ) ter primerljivost informacij obeh sistemov (koliko vsak sistem poda uporabnih informacij).

## Izpolnite vprašalnike SUS in SUISQ

Vprašalniki zajemajo dodatne informacije o uporabniku (vpisna št., spol, predhodna uporaba govornih sistemov), ki bodo služile za oblikovanje skupin uporabnikov in interpretacijo.

1. **Vprašalnik SUS za sistem A in še enkrat za sistem B**

[**https://docs.google.com/forms/d/e/1FAIpQLSfXUB\_vARHK0wb5pB\_khmUMzxOa4O2ZHkKBXGH5RZVt6-8C9w/viewform?usp=sf\_link**](https://docs.google.com/forms/d/e/1FAIpQLSfXUB_vARHK0wb5pB_khmUMzxOa4O2ZHkKBXGH5RZVt6-8C9w/viewform?usp=sf_link)

1. **Vprašalnik SUISQ-R za sistem A in sistem B**

<https://docs.google.com/forms/d/e/1FAIpQLSdUGdHCfmviRxI0MwpcTy-IKqwhTiEWkzqG05y1r9o1p_-kUQ/viewform>

A list of information on a computer

Description automatically generated with medium confidence

# Analiza podatkov odgovorov uporabnikov

Raziskovalna hipoteza:

Ali se kvaliteta dvogovora razlikuje med uporabo Aleksa in Google, in za katere faktorje skale?

Faktorji, ki jih izračunamo

* SUS - uporabnost
* UGO (User goal orientation) – uporabnikov cilj
* CSB (Customer service behavior) – obnašanje storitve
* SC (Speech characteristics) – značilnosti govora
* V (verbosity) - gostobesednost

Vstavite analizo podatkov in rezultate.

## Rezultati uporabnosti sistemov SUS

Izračunati moramo rezultat SUS (uporabnost storitve)

A screenshot of a test

Description automatically generated

Uvozimo podatke iz Excel datoteke v DataFrame

import pandas as pd

# Nalozi podatke iz lokalne datoteke

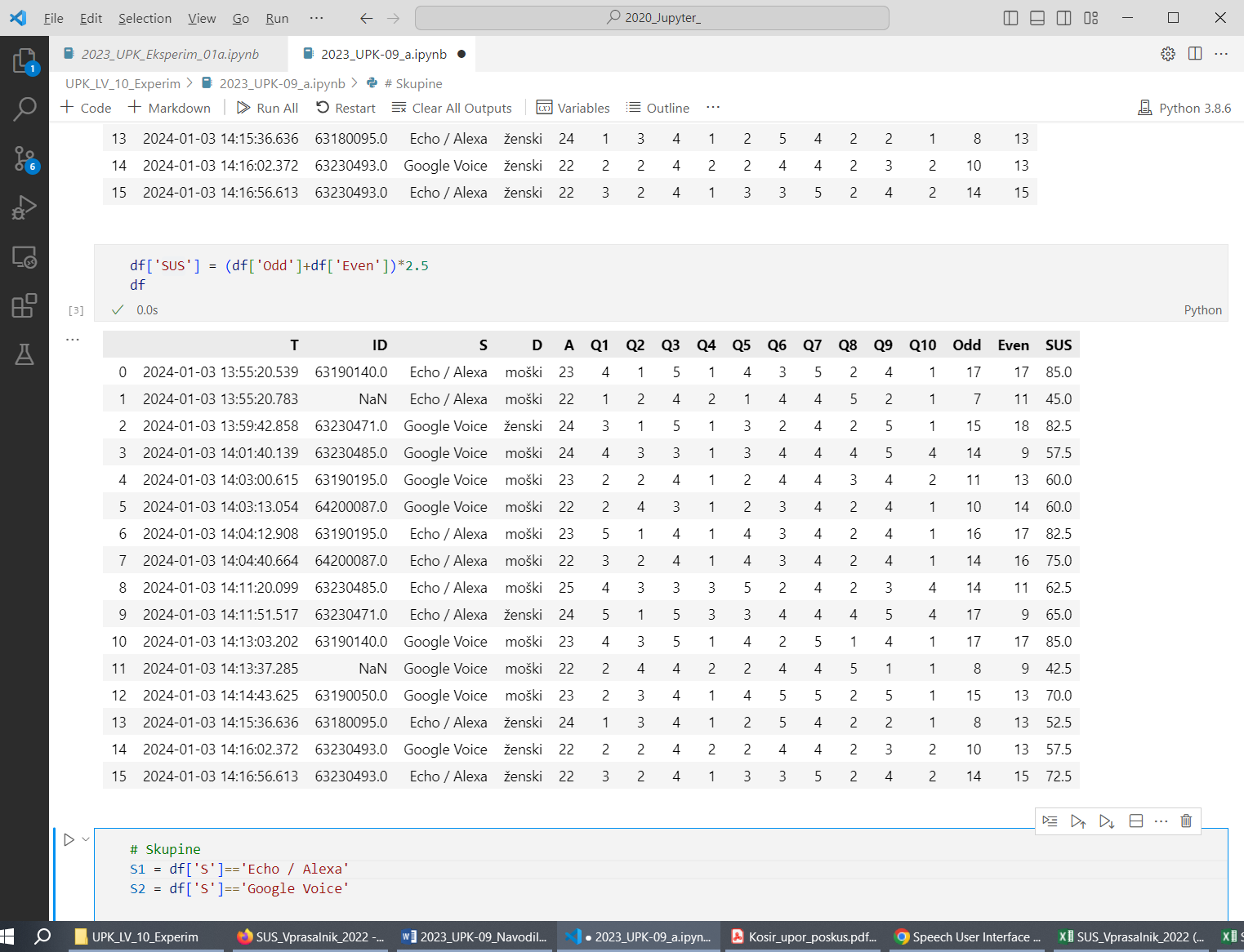
fn = 'SUS\_01a.xlsx'

df = pd.read\_excel(fn)

df

Po zgornji definiciji izračunamo stolpce Odd, Even in SUS

df['Odd'] = df['Q1']+df['Q3']+df['Q5']+df['Q7']+df['Q9'] - 5



Definiramo indekse skupin (Google, Alexa)

# Skupine

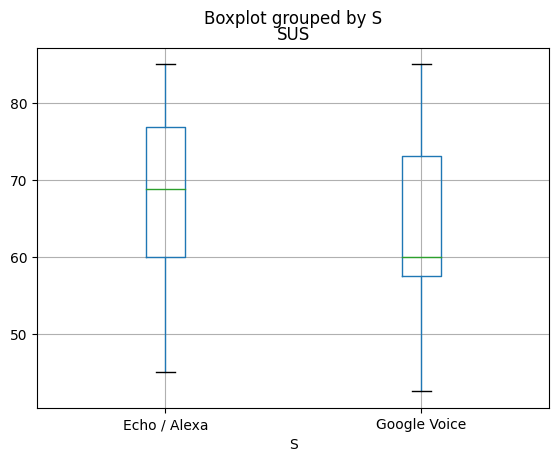
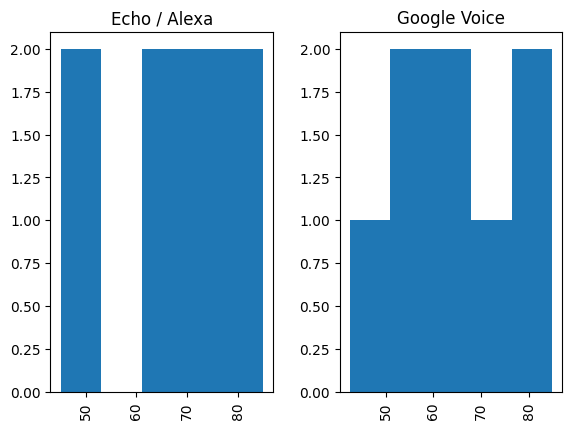
S1 = df['S']=='Echo / Alexa'

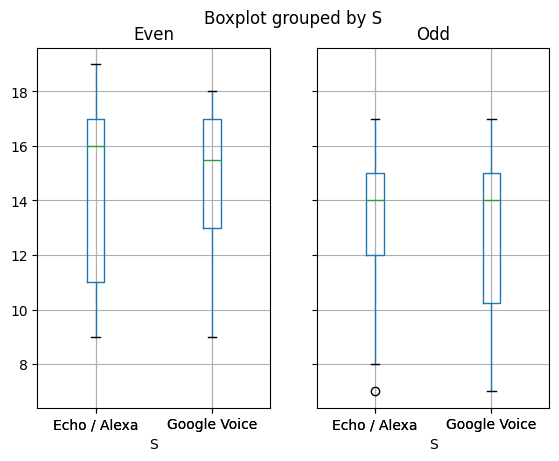
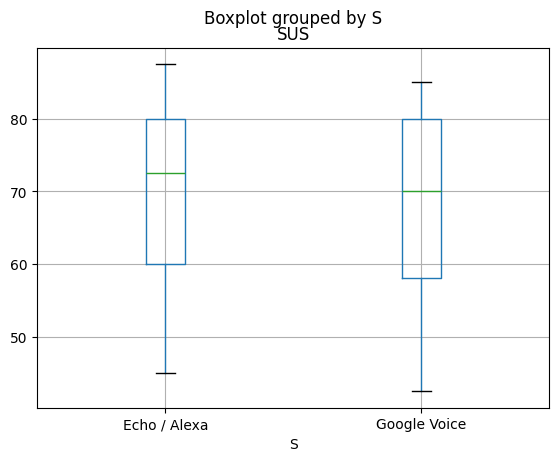
S2 = df['S']=='Google Voice'

Izrišemo rezultate po skupinah (boxplot)

df.boxplot(by='S', column=['Even','Odd'])

df.boxplot(by='S', column=['SUS'])



Preverimo s statističnim testom, ali je porazdelitev rezultatov normalna :

<https://www.statsmodels.org/devel/generated/statsmodels.stats.diagnostic.lilliefors.html>

# liliefors : Ali je porazdelitev normalna, p > 0.05

from statsmodels.api import stats

S1lil = stats.diagnostic.lilliefors(df[S1]['SUS'])

print("Normalna porazd p: ", S1lil[1])

S1lil

Vstavi rezultate, ali je normalna porazdelitev.

Interpretacija Rezultatov

. Normalna porazd p: 0.2235

· Rezultat testa: (0.168, 0.223)

· P-vrednost > 0.05:

predpostavimo, da je porazdelitev normalna.

· P-vrednost ≤ 0.05:

porazdelitev ni normalna.

**Preveri, ali je razlika med sistemi. Izračunaj p-faktor in interpretiraj rezultat.**

<https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.ttest_rel.html>

H0 : ničelna hipoteza, povprečja za oba sistema so enaka (ni razlike), (p>0.05)

Delovna hipoteza: povprečja se razlikujejo, med sistemi je razlika (p<0.05)

# TTest ničelna hipoteza: ni razlike med sistemi, p>0.05

from scipy import stats

stats.ttest\_rel(df[S1]['SUS'], df[S2]['SUS'])

rezultat, komentar

Normalna porazd p: 0.000999

(0.2953, 0.0009999)

TtestResult(statistic 0.686, pvalue=0.5033897645505774, df=14)

P-vrednost (pvalue): 0.5033897645505774. Ker je p-vrednost večja od 0.05, ni dovolj dokazov da sta si razlilčna za več kot nič. To pomeni, da ne moremo trditi, da obstaja statistično značilna

razlika med povprečji SUS ocen za sistema Echo/Alexa in Google Voice

Statistična vrednost (statistic): 0.6868606776955924

Ta vrednost predstavlja t-statistiko, ki se uporablja za izračun p-

vrednosti.

Stopnja svobode (df): 14

Število stopinj svobode, ki se uporablja pri izračunu t-statistike.

## Dodatna naloga: Izriši analize SUS

Uporabi spletno orodje za analizo SUS podatkov sistema A in sistema B:

<https://mixality.de/sus-analysis-toolkit/>

Pripraviti je potrebno csv datoteke za sistema Aleksa in Google in jih naložiti v orodje.

Rezultati, interpretacija:

(Sus reultat, primerjava z drugimi sistemi, ocena sistema, zanesljivost)

## Faktorji kvalitete pogovora SUISQ

Izračunaj faktorje:

1. UGO = User goal orientation: uporabnikov cilj

2. CSB = Customer Service Behaviors : obnašanje storitve

3. SC = Speech Characteristics: govorne lastnosti

4. V = Verbosity: gostobesednost, dolgoveznost

A list of information on a computer

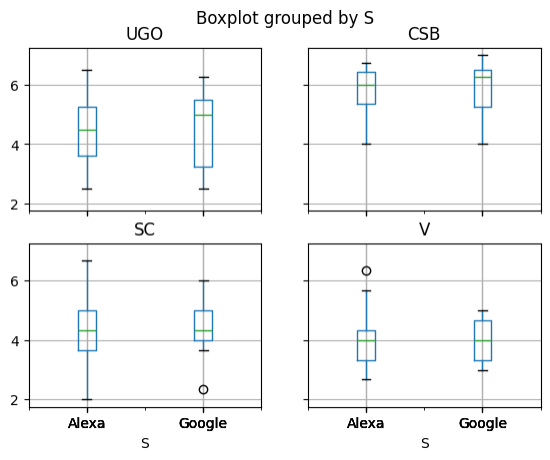
Description automatically generated with medium confidence

V\_r : obratni rezultat

Izračunaj po definiciji teh faktorjev:

Rezultat: izračunani faktorji (tabela)

UGO: TtestResult(statistic=-0.4617336564981453, pvalue=0.6513653305613312, df=14) CSB: TtestResult(statistic=-0.6793883090280631, pvalue=0.5079695759271889, df=14) SC: TtestResult(statistic=-1.053154786364301, pvalue=0.31010817331083324, df=14) V: TtestResult(statistic=0.3466578627692474, pvalue=0.7340035881011885, df=14)

Boxplot  


## Primerjaj sistema Aleksa, Google za 4 faktorje

Statistični T test za posamezne faktorje (ali velja ničelna ali delovna hipoteza):

Parameter UGO: User goal orientation: uporabnikov cilj

T-statistika: -0.4617

P-vrednost: 0.6514

Interpretacija: P-vrednost je večja od 0.05, zato ni statistično značilne razlike v uporabniški usmerjenosti med sistemoma Alexa in Google Voice.

Parameter CSB: Customer Service Behaviors

T-statistika: -0.6794

P-vrednost: 0.5080

Interpretacija: P-vrednost je večja od 0.05, zato ni statistično značilne razlike v obnašanju storitve med sistemoma.

Speech Characteristics (SC):

T-statistika: -1.0532

P-vrednost: 0.3101

Interpretacija: P-vrednost je večja od 0.05, zato ni statistično značilne razlike v govornih lastnostih med sistemoma.

Verbosity (V):

T-statistika: 0.3467

P-vrednost: 0.7340

Interpretacija: P-vrednost je večja od 0.05, zato ni statistično značilne razlike v gostobesednosti med sistemoma.

Boxplot faktorjev po sistemih (Aleksa, Google)