

Obligatorisk læringsaktivitet i Flow 1, Dataanalyse



Afleveringsfrist:

Onsdag den 9. oktober 2024, kl. 22

Udarbejdet af:

Thorbjørn Baum, adjunkt Cphbusiness

Thorbjørn Wulf, adjunkt Cphbusiness

Indholdsfortegnelse

Opgave 1 – Boliger og Danmarks Statistik	3
Opgave 1.1 – Det første skridt	3
Opgave 1.2 – Kategori-variabel	3
Opgave 1.3 – Merge de to dataframes	3
Opgave 1.4 – Plot	3
Opgave 2 – Forbrugertillidsindikatorer og fremtidig vækst i husholdningernes forbrugsudgift	4
Opgave 2.1 – Opdatering af DI's forbrugertillidsindikator	4
Opgave 2.2 – Forudsigelser af forbruget	4
Opgave 2.3 – Salg resten af året	4
Opgave 2.4 – Prognoser fra DI og Nationalbanken	4
Opgave 3 – Logistik regression, husholdningernes forbrugsudgift og forbrugertillid	5
Opgave 3.1 – Feature Engineering, dummy variable	5
Opgave 3.2 – Logistisk regression og forudsigelser	5
Opgave 3.3 – Simple validering af model	5
Opgave 3.4 – Potentielle forbedringer af model	5
Opgave 4 – Forbrug og forbrugertillidsindikatorer fra DST og DI, samt loops i lister	6
Opgave 4.1 – Illustration af forbrugertillid	6
Opgave 4.2 – Gennemsnit af underspørgsmål	6
Opgave 4.3 – De 11 grupper af forbrug	6
Opgave 4.4 – 22 simple lineære regressioner	6
Opgave 5 – Eurostat og API	7
Opgave 5.1 – Kvartalsvis årlig realvækst for en række Eurolande	7
Opgave 5.2 – Højeste kvartalsvise årlige realvækst	7
Opgave 5.3 – Coronakrisen som outlier	7
Opgave 5.4 – Effekt af Corona på forbruget	7

Opgave 1 – Boliger og Danmarks Statistik

Opgave 1.1 – Det første skridt

Skriv en kode, der viser hvordan du finder en tabel som kan give en liste over byer med indbyggertal vha DST-pakken dkstat.

Opgave 1.2 – Kategori-variabel

Lav en kategorivariabel i R ud fra følgende kriterier: `bycat = (c("landsby"=250, "lille by"=1000, "almindelig by"=2500, "større by"=10000, "storby"=50000))` og anvend den på listen over byer.

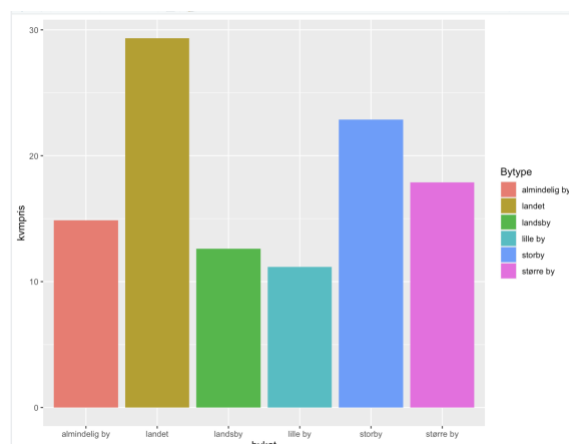
Opgave 1.3 – Merge de to dataframes

Indlæs filen med boliger og tilpas de to dataframes så du kan merge de to sammen via variabelen ”by” således at du får kategorien *bycat* med i dit bolig-datasæt.

Opgave 1.4 – Plot

Din merge skal producere en dataframe og et plot som vist nedenfor.

14	åbenrå	2.145.000 kr.	12.883	større by
15	åbenrå	2.495.000 kr.	12.600	større by
16	åbybro	1.595.000 kr.	8.886	almindelig by
17	åbybro	23.500.000 kr.	14.424	almindelig by
24	agerskov	3.995.000 kr.	5.022	lille by
25	åkirkeby	1.395.000 kr.	12.935	lille by
26	åkirkeby	1.095.000 kr.	9.144	lille by
27	åkirkeby	1.495.000 kr.	6.689	lille by
28	ålbæk	895.000 kr.	8.544	lille by
29	albertslund	3.695.000 kr.	28.314	større by



Opgave 2 – Forbrugertillidsindikatorer og fremtidig vækst i husholdningernes forbrugsudgift

Opgave 2.1 – Opdatering af DI's forbrugertillidsindikator

Opdatér DI's forbrugertillidsindikator med data frem til og med 2023 fra artiklen "Forbruget fortsætter fremgangen i 2016" (Baum, 2016). Lav vurdering af om forbrugertillidsindikatoren fra DI fortsat er bedre end forbrugertillidsindikatoren fra DST. (Hint: I bliver nødt til at nærlæse bilaget for DI-FTI for at finde starttidspunktet for estimationen, spørgsmålene, samt tabel, der sammenligner FTI og DI-FTI)

Opgave 2.2 – Forudsigelser af forbruget

Beregn/forudsig den årlige realvækst i husholdningernes forbrugsudgift for 3. kvartal 2024 med henholdsvis DI's forbrugertillidsindikator og forbrugertillidsindikatoren fra DST.

Opgave 2.3 – Salg resten af året

Med afsæt i jeres forudsigelse fra opgave 2.2, ville I så være bekymrede for virksomhedernes salg til hr. og fru Danmark resten af året. Giv en uddybende forklaring på jeres svar.

Opgave 2.4 – Prognoser fra DI og Nationalbanken

Hvor stor realvækst i privatforbruget forventer DI og Nationalbanken i deres seneste prognoser? Sammenhold deres tal med jeres svar i opgave 2.3.

Opgave 3 – Logistik regression, husholdningernes forbrugsudgift og forbrugertillid

Opgave 3.1 – Feature Engineering, dummy variable

Dan en dummy variable af den kvartalsvise årlige vækst i husholdningernes forbrugsudgift for perioden 1. kvartal 1998 til 2. kvartal 2021. Hvor ofte stiger den kvartalsvise årlige vækst i husholdningernes forbrugsudgift? Hvor ofte falder den kvartalsvise årlige vækst i husholdningernes forbrugsudgift?

Opgave 3.2 – Logistisk regression og forudsigelser

Lav en logistik regression med dummy variable fra opgave 1.1 og de fire indikatorer i DI's forbrugertillidsindikator. Hvilken retning forudsiger jeres model, at den årlige vækst i husholdningernes forbrugsudgift, vil gå i 3. kvartal 2024? (Hint: svaret er enten op eller ned)

Opgave 3.3 – Simpel validering af model

Hvor ofte forudsiger jeres model i opgave 1.2, at den kvartalsvise årlige realvækst i husholdningernes forbrugsudgift stiger? Hvor ofte er det så reelt tilfældet, at den kvartalsvise årlige realvækst i husholdningernes forbrugsudgift stiger, set i forhold til, hvad jeres model forudsiger? (hint: hvor mange tilfælde af 1 og 1 finder jeres model?)

Opgave 3.4 – Potentielle forbedringer af model

Hvordan I prøve at forbedre jeres model? Lav 2 valgfrie scenarier og test dem i forhold til jeres baseline (underspørgsmålene fra DI's forbrugertillidsindikator).

Opgave 4 – Forbrug og forbrugertillidsindikatorer fra DST og DI, samt loops i lister

Opgave 4.1 – Illustration af forbrugertillid

Hent data for forbrugertillidsundersøgelsen fra januar 1996 til i dag og omregn jeres data til kvartaler. Lav en grafisk illustration af jeres omregnede data for DST's forbrugertillidsindikator og kommentér på, hvornår de danske forbrugere er mest og mindst optimistiske.

Opgave 4.2 – Gennemsnit af underspørgsmål

Beregn gennemsnittet for underspørgsmålet "Set i lyset af den økonomiske situation, mener du, at det for øjeblikket er fordelagtigt at anskaffe større forbrugsgoder som fjernsyn, vaskemaskine eller lignende, eller er det bedre at vente?" for perioden 1. kvartal 2000 til og med 3. kvartal 2024. Vurdér jeres resultat set i forhold til spørgsmålet og svarmulighederne. (Hint: giver resultatet analytisk mening?)

Opgave 4.3 – De 11 grupper af forbrug

Hent data for de 11 grupper af forbrug blandt husholdningerne. Hvad brugte danskerne flest penge på i 2023? Hvilken gruppe af forbruget steg mest fra 2020 til 2023? (hint: I kan ikke lægge kvartalerne sammen, når I har kædede værdier)

Opgave 4.4 – 22 simple lineære regressioner

Lav 22 simple lineære regressioner mellem hver af de 11 grupper i forbruget (y-variable) og henholdsvis forbrugertillidsindikatoren fra DST og DI. I skal gemme summary i 22 lister. I skal lave jeres regressioner fra 1. kvartal 2000 til og med 2. kvartal 2024.

Opgave 5 – Eurostat og API

Opgave 5.1 – Kvartalsvis årlig realvækst for en række Eurolande

Beregn den kvartalsvise årlige realvækst for husholdningernes forbrugsudgift for Danmark, Belgien, Holland, Sverige, Østrig, Tyskland, Frankrig, Italien og Spanien i perioden 1. kvartal 2000 til og med 2. kvartal 2024. I skal hente data vha. API'et fra Eurostat.

Opgave 5.2 – Højeste kvartalsvise årlige realvækst

Hvilket af de landene har gennemsnitligt haft den højeste kvartalsvise årlige realvækst i husholdningernes forbrugsudgift i perioden 1. kvartal 2000 til 2. kvartal 2024.

Opgave 5.3 – Coronakrisen som outlier

Fjerne Coronakrisen fra jeres data og find igen den gennemsnitligt kvartalsvise realvækst i husholdningernes forbrugsudgift i perioden 1. kvartal 2000 til 2. kvartal 2024. I hvilket af landene har Coronakrisen haft en største effekt på den gennemsnitligt kvartalsvise realvækst.

Opgave 5.4 – Effekt af Corona på forbruget

I hvilket europæiske land faldt den gennemsnitligt kvartalsvise realvækst i husholdningernes forbrugsudgift, i perioden 1. kvartal 2020 til 2. kvartal 2024, mest?