1. Vem du har arbetat i grupp med?

I min grupp har jag jobbat med David Norman, Erik Fredrixon, Filip Holmberg, Magdalena Wallner och Robert Westman.

1. Hur har ni i gruppen arbetat tillsammans?

Generellt så tycker jag att det har gått bra och vi har vart duktiga på att lyssna på varandra i de olika val vi gjort. Även om jag inte håller med i alla beslut så har jag respekterat gruppens önskan och utför de olika delar vi kommit överens om.

1. Vad var bra i grupparbetet och vad kan utvecklas?

Väldigt bra kommunikation och vilja att hjälpa varandra. Vissa i gruppen ville verkligen göra en webbskrapare vilket tog ganska mycket tid och blev inte klar för än de flesta redan hade samlat sin data manuellt. Detta tyckte jag var onödigt och tidskrävande.

1. Vad är dina styrkor och utvecklingsmöjligheter när du arbetar i grupp?

Bli bättre på att prata i grupp, jag har alltid tyckt det är obehagligt att prata i grupper som är större än tre personer.

1. Finns det något du hade gjort annorlunda? Vad i sådana fall?

Inte tagit med färg och istället för att vi tog ett märke var så ha helt slumpmässigt vilka bilar vi tog. Inte haft ett prisspann.

Teoretiska frågor

1. Kolla på följande video: https://www.youtube.com/watch?v=X9\_ISJ0YpGw&t=290s , beskriv kortfattat vad en Quantile-Quantile (QQ) plot är.

QQ-plot är ett effektivt sätt för att se om ens data är normalfördelad.

1. Din kollega Karin frågar dig följande: ”Jag har hört att i Maskininlärning så är fokus på prediktioner medan man i statistisk regressionsanalys kan göra såväl prediktioner som statistisk inferens. Vad menas med det, kan du ge några exempel?” Vad svarar du Karin?

När man exempelvis försöker prediktera värdet på en bil så finns det flera olika parametrar som påverkar värdet. I maskininlärning så bryr man sig inte särskilt mycket hur dessa parametrar påverkar medan i statistiks inferens så gör man det. Alltså så bryr man sig hur en modell påverkas av de olika parametrar man har med i modellen.

1. Vad är skillnaden på ”konfidensintervall” och ”prediktionsintervall” för predikterade värden?

Konfidensintervall är en intervallskattning som ger en uppskattning på ett parametervärde. Om exempelvis ett 95% konfidensintervall för medelvärdet av en population är 1 till 5 så är vi 95% säkra att det verkliga medelvärdet är mellan 1 och 5.

Prediktionsintervall är en intervallskattning som get osäkerheten i en prognos för en framtida observation. Till skillnad från konfidensintervallet så tar prediktionsintervallet hänsyn till den slumpmässiga variationen i observationerna och modellens osäkerhet.

1. Den multipla linjära regressionsmodellen kan skrivas som: 𝑌 = 𝛽0 + 𝛽1𝑥1 + 𝛽2𝑥2+ . . . + 𝛽𝑝𝑥𝑝 + 𝜀 . Hur tolkas beta parametrarna?

För varje en enhet, exempelvis, x1 ökar så ökar Y med 𝛽1 då alla andra parametrar är konstanta.

1. Din kollega Hassan frågar dig följande: ”Stämmer det att man i statistisk regressionsmodellering inte behöver använda träning, validering och test set om man nyttjar mått såsom BIC? Vad är logiken bakom detta?” Vad svarar du Hassan?

BIC används generellt för att testa olika modeller mot varandra, inte för att se hur en enskild modell presterar. För att använda BIC krävs inte test set men om du vill testa modellens generaliseringsförmåga på ny data så ska du använda dig av ett test set.

1. Förklara algoritmen nedan för ”Best subset selection”

En bild som visar text, skärmbild, Teckensnitt, nummer

Automatiskt genererad beskrivning

Först skapas en modell med noll prediktorer. Sedan söker best subset selection algoritmen efter den prediktor som presterar bäst själv. Sedan söker man efter de två prediktorer som presterar bäst tillsammans osv tills man har en modell för alla antal prediktorer. Detta kan ta väldigt lång tid då algoritmen går igenom varje kombination av prediktorer.

1. Ett citat från statistikern George Box är: “All models are wrong, some are useful.” Förklara vad som menas med det citatet.

Då en modell i princip aldrig ger dig rätt svar så kan den ge en utbildad gissning som är tillräckligt nära det rätta svaret att det kan vara användbart.

Självutvärdering

1. Utmaningar du haft under arbetet samt hur du hanterat dem.   
   Att förstå alla värden så som BIC, Cp, MAE. Det var varit väldigt mycket nya värden på kort tid. Ta en sak i taget, experimentera och se hur värdet ändras om man gör olika ändringar. Sen har det hjälp att kolla på youtube för lite exempel eller fråga chatgpt för definitioner.
2. Vilket betyg du anser att du skall ha och varför.   
   VG, jag tycker att jag uppfyller kraven. Det jag gjort har jag gjort på ett kontrollerat sätt och har förståelse för.
3. Något du vill lyfta fram till Antonio?

Inget om kursen. Men om vi inte ses mer innan lovet så får du ha en trevlig sommar och vi ses nästa termin!

En bild som visar text, skärmbild, linje, Graf

Automatiskt genererad beskrivning

Figur 1 Antal nyregistrerade bilar mellan 2014 och 2019

## Nyregistrerade Bilar i Västra Götaland

Innan 2020 kan man se att antalet nyregistrerade bilar är relativt konstant utöver en mindre sjunkning runt sommaren 2018. Medan tittar man under och efter 2020 så ser man en relativt kraftig minskning i nyregistrerade bilar. Detta är antagligen pga. corona-viruset som bröt ut i mars 2020 följt av kriget i Ukraina som började i februari 2022 och den kraftiga ökningen i inflation man sätt som direkt påföljd av kriget.

En bild som visar text, skärmbild, linje, Graf

Automatiskt genererad beskrivning

Figur 2 Antal nyregistrerade bilar mellan 2020 och 2024