

**Framsida**

# **Examensarbete**

## **Försäljningsprognoser för elektronisk butik**

Khalid Youssouf Handouleh

Data Scientist

Examensarbete, 30 yhp

Handledare: Carl Jendle

Göteborg 2022-02-07

## Innehållsförteckning

### Contents

Framsida.....	1
Innehållsförteckning .....	2
Bakgrund, syfte och frågeställning.....	3
Avgränsningar .....	4
Metod .....	4
Teoretiskt arbete: .....	4
Praktiskt arbete:.....	5
Time Series Analysis.....	7
Resultat.....	8
Praktiskt arbete:.....	8
Teoretiskt arbete: .....	8
Prognosvalidering .....	9
Diskussion .....	10
Källförteckning .....	10

## Bakgrund, syfte och frågeställning

Förutsägelse av försäljning är avgörande eftersom det kan hjälpa företaget att urskilja vad som går rätt, samt vilka zoner i dess nuvarande metodik som bör justeras och ändras för att garantera framtida framgång. För fall, om säljgruppen är tillförlitligt under standarder, kan försäljningsförutsägelse erbjuda hjälp att avgöra i vilken region och vilken kategori dessa problem händer.

Dessutom hjälper det att veta hur försäljningen förändras beroende på månader och olika säsonger. Prognoser kan dessutom hjälpa till att välja framtida handelsträvanden och ger företagets förmåga att fatta välgrundat affärsbeslut och utveckla datadrivna strategier.

Som Data Scientist Student med affärsbakgrund gynnar detta mig när det kommer till att lära mig mer om produkt- och försäljningsanalyser samt att prognostisera framtida försäljning med hjälp av prediktiv analysverktyg som Python och Business Intelligence.

I det här projektet kommer jag att förutsäga försäljningen för Electronic Store.

Den historiska informationen innehåller mängder av register för försäljning av elektroniska produkter mellan jan-2016 och dec-2019. Butikerna finns i olika regioner och varje butik innehåller flera segment. Butiken säljer dessa tre huvudsakliga elektroniska kategorier: kameror, bärbara datorer och mobiltelefoner som säljs av dessa tre kundsegment: konsument, företag och hemmakontor.

” Länk till dataset: [https://data-flair.training/blogs/download-electronic-store-sales-data/?ref=morioh.com&utm\\_source=morioh.com](https://data-flair.training/blogs/download-electronic-store-sales-data/?ref=morioh.com&utm_source=morioh.com)

Först och främst ska jag analysera data generellt och svara på dessa frågor.

Vem sålde mest (Kundsegment)?

Vilken produktkategori såldes mest?

Vilken stat sålde mest kvantitet?

Vilken produktkategori såldes mest och i vilket stat?

fem stater som har störst försäljning?

Jag ska också:

Förutsäg hur mycket det totala försäljningsbeloppet kommer att vara under en säsong.

Totala försäljningsbeloppet under en månad enligt försäljningen under 2019.

Förutsäg hur mycket det totala försäljningsbeloppet blir för varje kategori under en säsong enligt försäljningen 2019

## Avgränsningar

Rapporten är begränsad till att förutsäga månads-/säsongsförsäljning inom python-maskininlärning. Jag valde att använda övervakad maskininlärning eftersom För att kunna prognostisera försäljningen för en viss produktkategori under en specifik tidsperiod och i en viss region - till exempel försäljning av kameror för vintersäsongen i Västra regionen finns det ett behov av att tänka produktkategoriförändringar förutom tidsserien.

## Metod

I det här projektet använde jag statistiska analysfunktioner i pandas och visualisering från matplotlib och seaborn med målet att upptäcka användbar information och svara på frågor.

För att förutsäga försäljning använde jag supervised machine learning algorithms.

## Teoretiskt arbete:

Pandas statistiska analysfunktioner och visualisering från matplotlib och seaborn är lätt för att hitta användbar information genom att skriva en enkel kod. Därför kommer jag att använda dem för att visualisera data och svara på dessa frågor.

Vem sålde mest (Kundsegment)?

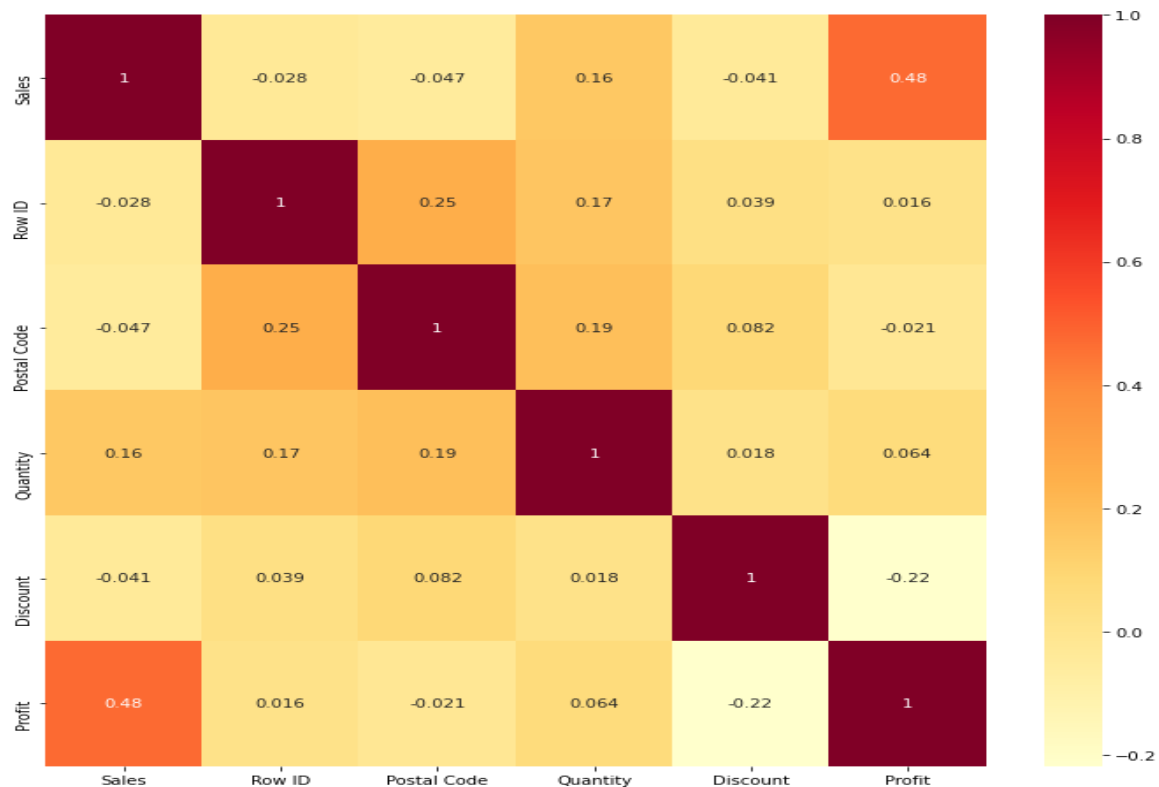
Vilken produktkategori såldes mest?

Vilken stat sålde mest kvantitet?

Vilken region har det högsta totala försäljningsbeloppet?

Denna data innehåller 9994 observationer och 21 variabler och har inga saknade värden.

För att förutsäga kvartalsförsäljning för hela datan kommer jag att gruppera datumkolumnen till kvartalsförsäljning genom att använda pandas och beräkna korrelationer mellan variabler. Med matplotlib visualiserade jag denna korrelation i följande heatmap.



**Figur 1.** Correlation\_Heatmap\_Plot

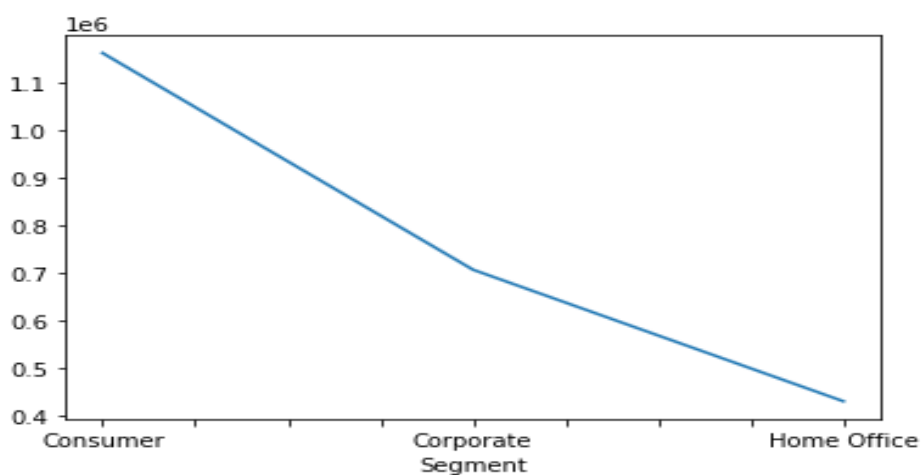
Från figur 1, ser vi att försäljning och vinst har störst korrelation. Försäljning och kvantitet är också något korrelerade med varandra. Därför använder jag delmängd av data genom att välja Date, Quantity, Profit and Sales.

## Praktiskt arbete:

### Vem sålde mest (Kundsegment)?

```
segment_sales= df.groupby(["Segment"]).sum().sort_values("Sales", ascending=False)
```

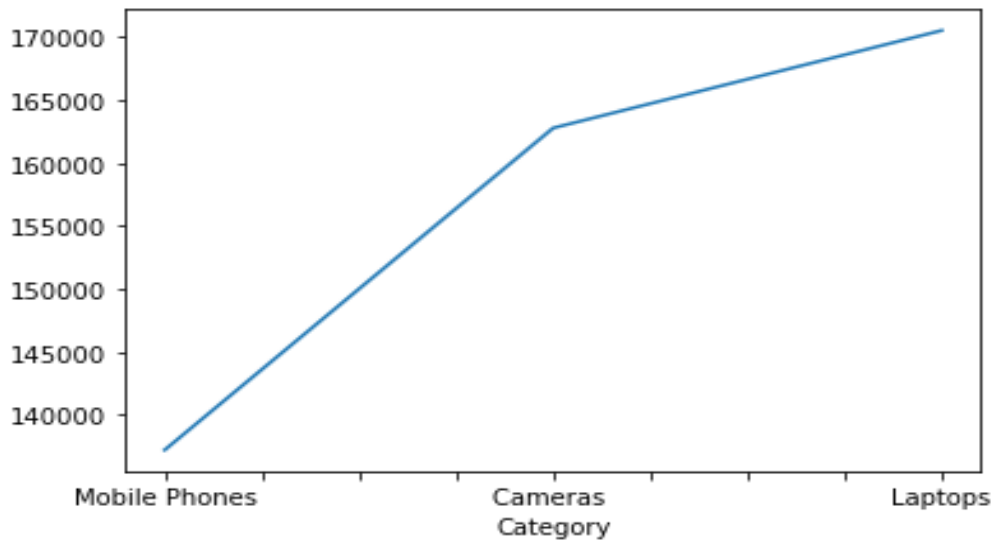
```
segment_sales['Sales'].plot()
```



**Figur 2.** Segmetn\_Sales\_Plot

### Vilken produktkategori såldes mest?

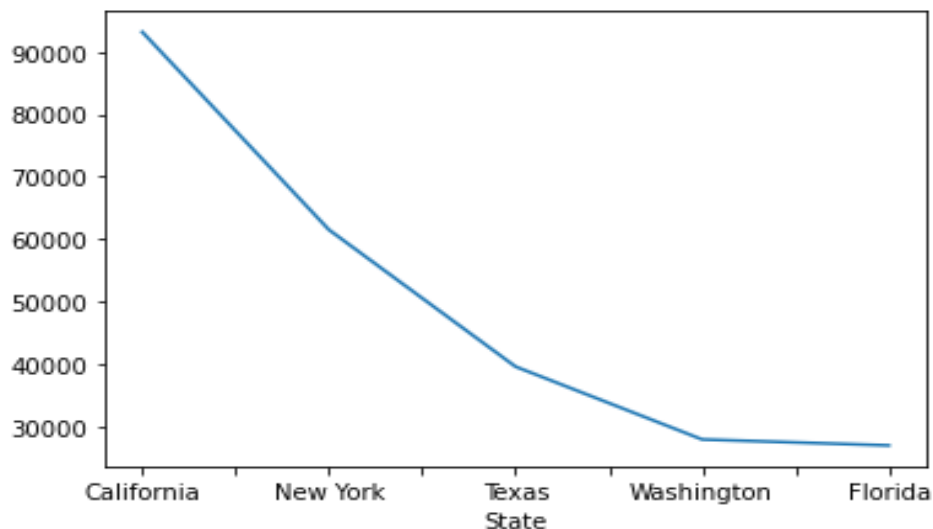
```
category_sales= df.groupby(["Category"]).sum().sort_values("Sales", ascending=True).  
category_sales['Sales'].plot()
```



**Figur 3.** Category\_Sales\_Plot

### Vilken stat sålde mest kvantitet?

```
group = df.groupby(["State","Quantity"]).sum()  
total_sales_Quantity= group["Sales"].groupby(level=0, group_keys=True)  
total_sales_Quantity.max().nlargest(5).plot()
```

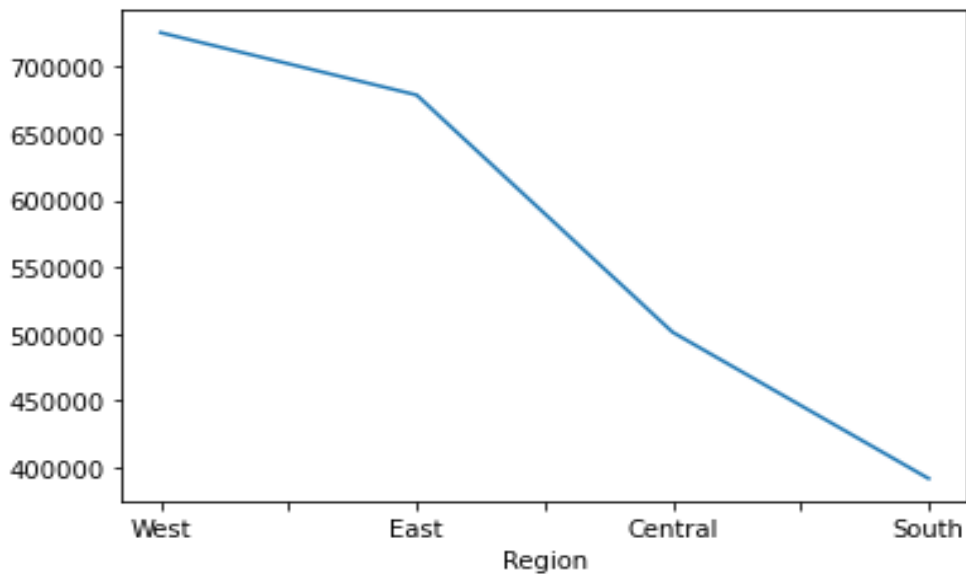


**Figur 4.** State\_Sales\_Plot

### Vilken region har det högsta totala försäljningsbeloppet?

```
region_sales= df.groupby(["Region"]).sum().sort_values("Sales", ascending=False)
```

```
region_sales['Sales'].plot()
```



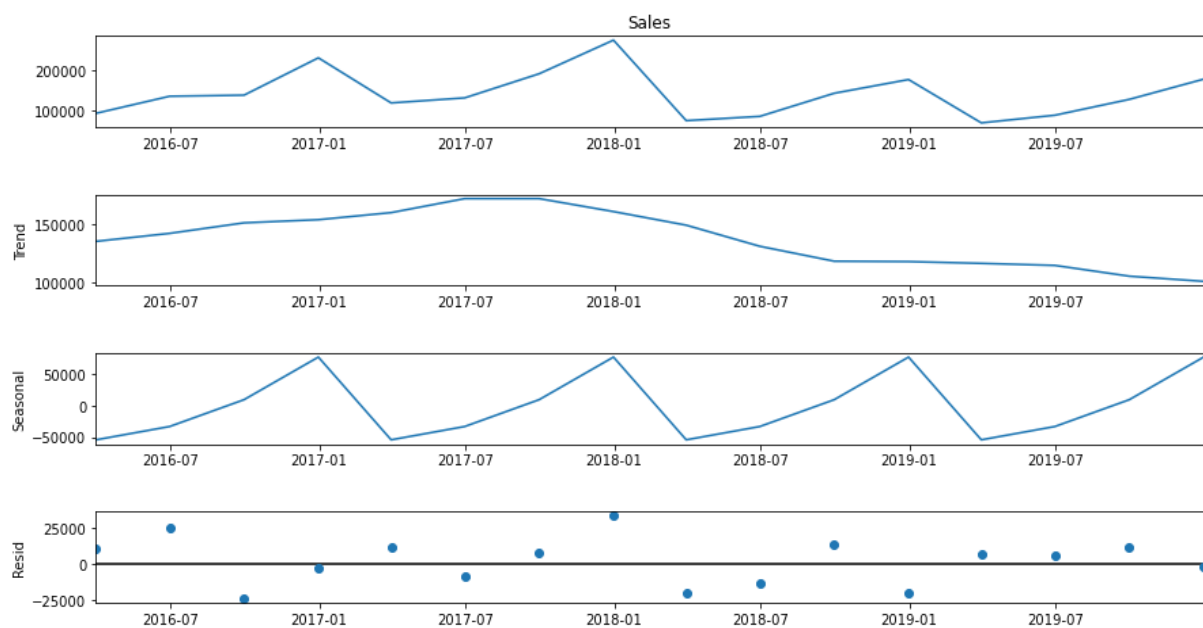
**Figur 5.** Region\_Sales\_Plot

Efter att ha plottat data ser vi att försäljningen ändras kvartalsvis uppåt från början till slutet av året.

## Time Series Analysis

I tidsserieanalys kommer jag att dela upp en tidsserie i tre komponenter: en trendcykelkomponent, en säsongskomponent och en restkomponent.

Följande tidsserieuppdelning visar datas säsongsvariation.



**Figur 5.** Time series decomposition

Därför kommer jag att utföra kvartalsvis försäljningsprognoser för produktkategorier genom att använda ARIMA-modellen. För att förutsäga den totala försäljningen kvartalsvis kommer jag att använda övervakad maskininlärning. I slutet kommer jag att förutsäga kvartalsförsäljning i kategorin i en specifik region.

## Resultat

Som figur 2 visar är kund det konsumentsegment som sålt flest produkter medan den största produktkategoriförsäljningen som resulterade är Mobiltelefoner som figur 3 illustrerar. När det kommer till vilken delstat som sålde mest fann jag att Kalifornien kommer först som figur 4 visar. Vi ser också i figur 5 att region Väst har de största försäljningsenheterna.

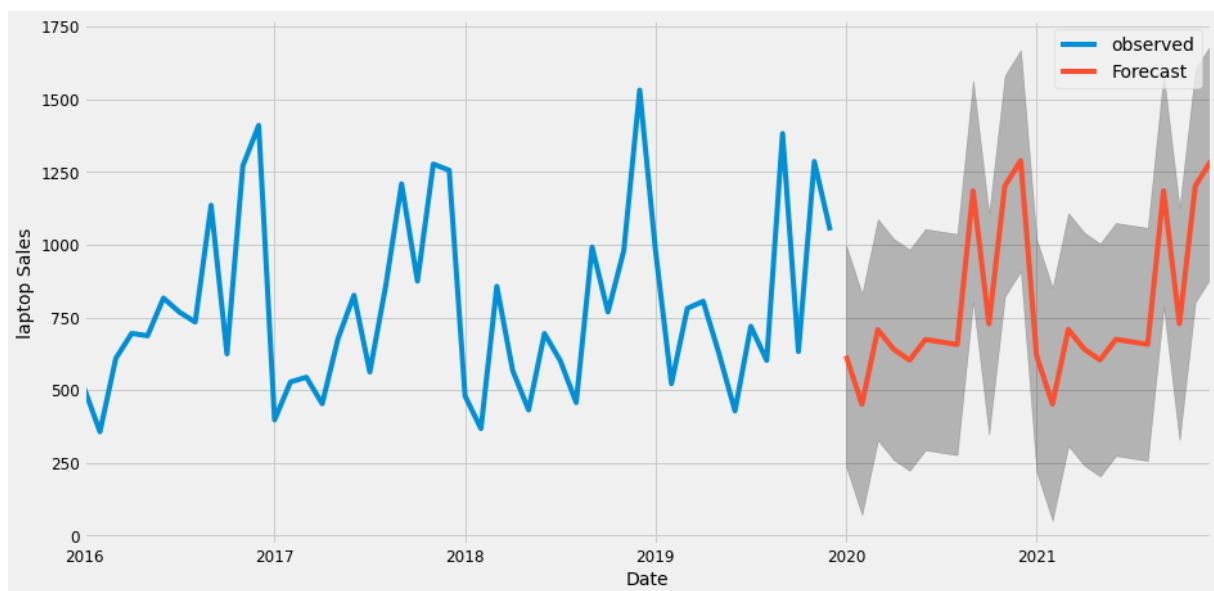
## Praktiskt arbete:

Länk till dataset: [https://data-flair.training/blogs/download-electronic-store-sales-data/?ref=morioh.com&utm\\_source=morioh.com](https://data-flair.training/blogs/download-electronic-store-sales-data/?ref=morioh.com&utm_source=morioh.com)

Projektkoder laddas upp som separata filer i ping-pong

## Teoretiskt arbete:

Vår modell fångade tydligt säsongvariationerna i försäljningen av bärbara datorer. Som vi förutspår längre in i framtiden. Detta återspeglas i de konfidensintervall som genereras av vår modell, som växer sig större när vi går längre från början till slutet av varje år. Figur 6 illustrerar tidsserieprognos för resultat för försäljning av bärbara datorer.

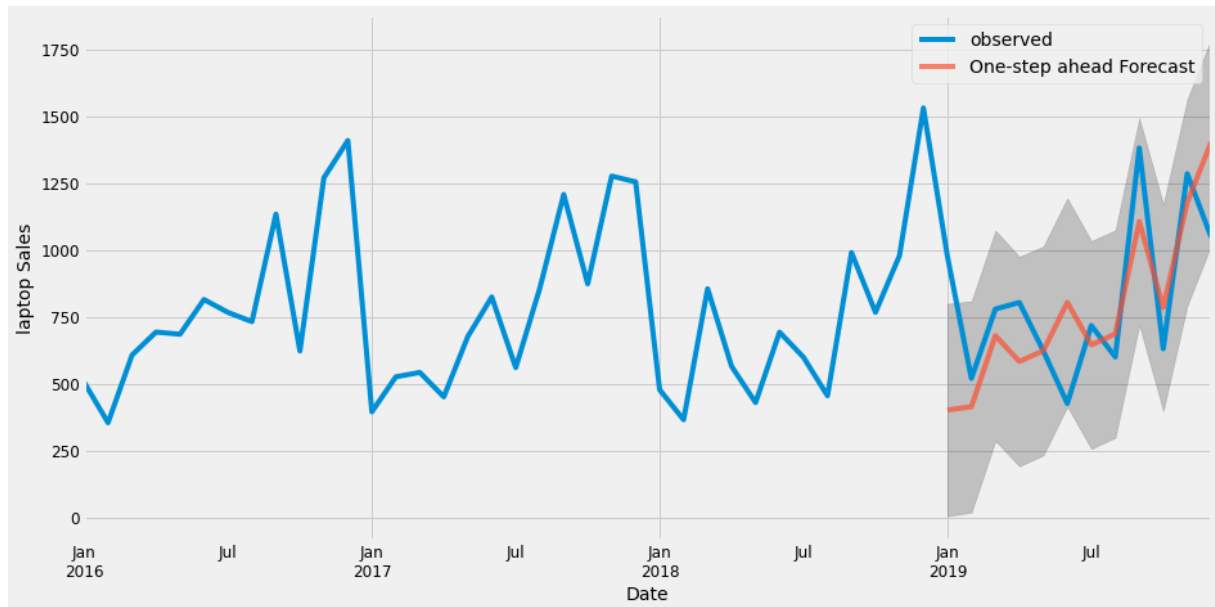


**Figur 6.** Prognos\_Results



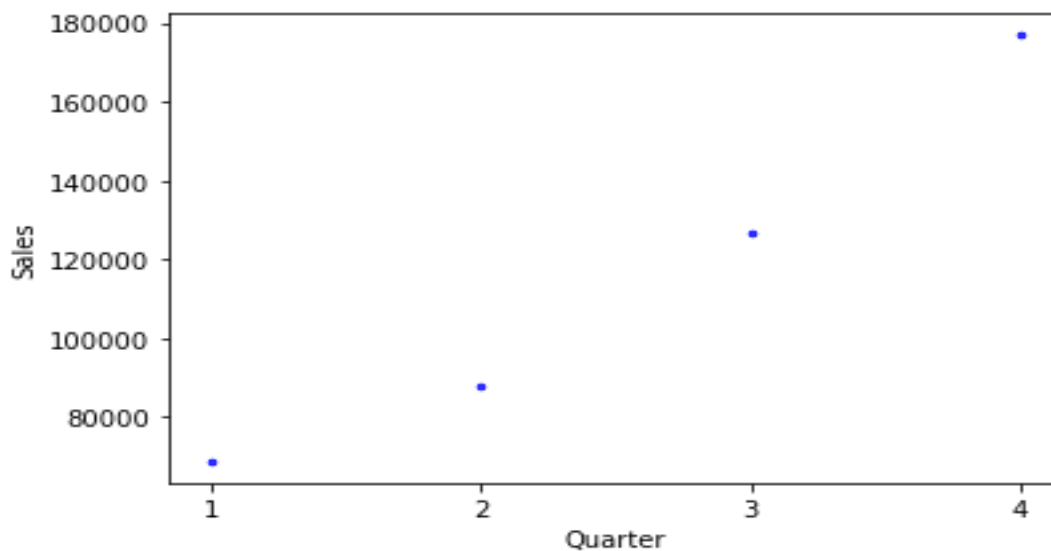
## Prognosvalidering

För att validera prognoser jämför vi förväntad försäljning med faktisk försäljning av tidsserien, och vi sätter prognoser att börja 2019-01-01 fram till slutet av data. Linjediagrammet i figur 7 visar de observerade värdena jämfört med de rullande prognoseerna.



**Figur 7.** Prognosvalidering

För att förutsäga den totala försäljningen för nästa kvartal använde jag linear regression model för försäljningsdata under 2019 eftersom vi kan se i figur 8 att försäljningen är linjär.



**Figur 8.** Quarter\_Sales

Med att betrakta kvartal från 1 till 3 som träningsdata och kvartal 4 som testdata, förutspår Ride Model 180589,3 som den totala försäljningen nästa kvartal medan det faktiska antalet produktförsäljningar är 176913,3

Actual\_Sales Predicted\_Sales

176913.2982 180589.264179

Förutsägelsen av den månatliga kategoriförsäljningen för datamängden Region\_Prod\_Total\_Sales.xlsx visar den totala kategoriförsäljningen för oktober, november och decembermånaderna 2019.

Actual\_Sales Predicted\_Sales

31404.9235 31404.960158

75972.5635 75972.717811

74919.5212 74920.092182

Mer djupgående utförde jag förutsägelsen av en specifik kategoriförsäljning under ett kvartal och i specifik region. Till exempel förutspådde jag att kameraförsäljningen för Västregionen under första kvartalet 2019 var 9622,8 medan den faktiska försäljningen var 9178,6.

## Diskussion

I detta projekt försökte jag täcka försäljningen av elektronisk butik och svara på frågor. Jag utförde tidsserieanalyser för att förutsäga hur produktkategoriförsäljningen kommer att förändras baserat på årets kvartal, men detta räckte inte, eftersom jag behövde prognostisera hur mycket den totala försäljningen kommer att bli under nästa månad eller säsong. Därför använde jag övervakade maskininlärningsalgoritmer och försäljning av 2019. Mer djupgående utförde jag förutsägelsen av en specifik kategoriförsäljning under ett kvartal och i specifik region.

För att lösa problemet och svara på alla frågor behövde jag göra fler pythonfiler och utvalda delmängder av datan. Under resan lärde jag mig mer om datamanipulation och tidsserieanalys, vilket var första gången för mig att använda det.

## Källförteckning

### Webbkällor

<https://medium.com/analytics-vidhya/statistical-analysis-in-python-using-pandas-27c6a4209de2>

<https://www.atmosera.com/blog/using-pandas-to-analyze-sales-data/>

<https://www.bounteous.com/insights/2020/09/15/forecasting-time-series-model-using-python-part-one/>

[An End-to-End Project on Time Series Analysis and Forecasting with Python | by Susan Li | Towards Data Science](#)

<https://medium.com/analytics-vidhya/python-code-on-holt-winters-forecasting-3843808a9873>