

## PRESENTATION SMOKING HABITS

Here we have an example how certain transactions Data can be used to help understand Patiens smoking habits. Transactions can be viewed by using either BankID or related Bank account transactions statements (lists and tables).

In this example we are logging into an Nordea Bank account and selecting transactions for last 3 months.

The screenshot shows the Nordea Internetbanken Privat - Förenklad interface. The top navigation bar includes links for Vardagsärenden, Spara och placera, Lån, Skadeförsäkringar, Information och råd, and Inställningar. The left sidebar lists various services like Betala och överföra, Konton, Kort, Mobila tjänster, Avtal och meddelanden, Dokument, and BankID. The main content area is titled 'Kontoutdrag och kontoinformation' and features two tabs: 'Kontoutdrag' (selected) and 'Kontoinformation'. Under the 'Kontoutdrag' tab, there are fields for Konto, Saldo, Att disponera, and Datum. A dropdown menu for 'Senaste 3 månaderna' is visible next to the Datum field.

At the bottom of the transactions list I am offered to save the list as Excel or Csv file.

<input type="checkbox"/>	2019-03-06	<a href="#">Kortköp 190304 HEMKÖP STOCKHOLM GUL</a>
<input type="checkbox"/>	2019-03-04	<a href="#">MARTINA.DOM</a>
<input type="checkbox"/>	2019-03-04	<a href="#">Kortköp 190228 HEMKÖP STOCKHOLM GUL</a>
<input type="checkbox"/>	2019-03-01	<a href="#">Vardagspaket Tillval Betalningar</a>
<input type="checkbox"/>	2019-03-01	<a href="#">Vardagspaket Månadspris kort</a>

Spara som fil: [Excel](#) [CSV](#)

## PRESENTATION SMOKING HABITS

With Python I can convert that table into a DataFrame and look for specific values. (For the purpose of this presentation I have already dropped the SALDO category and am now trying to search for specific transactions of 66SEK - that is the cost of a single Marlboro Gold pack.)

```
df.loc[df['Belopp'] == '-66,00']
```

	Datum	Transaktion	Kategori	Belopp
15	2019-04-15	Kortköp 190413 7 ELEVEN 4119113	NaN	-66,00
29	2019-04-08	Kortköp 190405 DIREKTEN TAREQS TOBA	NaN	-66,00
46	2019-04-08	Kortköp 190405 SHML KIOSK	NaN	-66,00
70	2019-03-11	Kortköp 190309 GOTGATSBACKEN	NaN	-66,00

In this displayed result we can observe that 4 packs of Marlboro were bought since March 1st. What is interesting is that 2 were bought on the same date.

If I examine all costs on that particular date I can find out some repetitive transactions which would usually mean drinks or rounds of drinks at a bar. In this case 3 rounds of drinks were bought at that particular date.

```
df.loc[df['Datum'] == '2019-04-08']
```

	Datum	Transaktion	Kategori	Belopp
32	2019-04-08	Kortköp 190405 GOTGATSBACKEN	NaN	-101,00
33	2019-04-08	Kortköp 190405 GOTGATSBACKEN	NaN	-101,00
34	2019-04-08	Kortköp 190405 GOTGATSBACKEN	NaN	-101,00

What does this all mean ? In this case we can conclude that the 'Patient' is a light smoker who tends to smoke over a drink. Where would that classify him as a smoker when determining the Risk Factor at the COXA clinic? SMOKER STATUS category could be expanded from YES and NO to: No, Light, Medium, Heavy and at later point we could transform this Categorical values to Numerical (0,1,2,3) which would make it much more easier and more precise when using Simple Linear Regression to determine the overall Risk Factor for the Patient.