Hi Guys,

Here is my analysis :

1. For brand 5 , 6 and 8 people are buying more expensive packets as compared to the cheaper ones.

Price_per_Packet \ Brand	1	2	3	4	5	6	7	8	9	Totals
1	345	1,163	4,915	7,359	912	1,922	1,393	2,553	9,075	**29,637**
2	12,030	5,843	8,917	2,011	2,627	4,302	3,496	8,568	19,580	**67,374**
3	3,287	1,866			2,823	7,050	3,122	8,934	6,893	**33,975**
Totals	**15,662**	**8,872**	**13,832**	**9,370**	**6,362**	**13,274**	**8,011**	**20,055**	**35,548**	**130,986**

2. Most people coming to buy again or “Days between purchase” is highest for 7 days which is 16369 where probability is 12.49% but the median is 9 which means most of the people buy again after 9 days and standard deviation is 20.58. The value of SD indicates that how much data is spread out from the average.

```
> summary(Days_between_Purchase)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   1.00    6.00    9.00   15.14   17.00  741.00
> names(which.max(table(Days_between_Purchase)))
[1] "7"
> sd(Days_between_Purchase)
[1] 20.58186
```

3. People belonging to “household size 2” buys the most “1 packet” which is 37869.

No_of_Packet \ Household_Sz	1	2	3	4	5	Totals
1	19,069	37,869	22,132	16,175	9,777	**105,022**
2	1,983	7,785	4,363	3,013	2,145	**19,289**
3	535	2,970	1,415	1,012	743	**6,675**
Totals	**21,587**	**48,624**	**27,910**	**20,200**	**12,665**	**130,986**

4. People belonging to “household size 2” consumes more coffee than anybody and also buys the most famous brand “Andere Kaffeemarken” which is 13825.

Brand \ Household_Sz	1	2	3	4	5	Totals
1	3,039	6,031	3,195	1,905	1,492	**15,662**
2	1,611	3,567	1,945	1,061	688	**8,872**
3	2,256	5,243	2,685	2,471	1,177	**13,832**
4	1,048	3,502	2,400	1,596	824	**9,370**
5	974	2,164	1,257	1,088	879	**6,362**
6	2,176	4,974	2,837	2,228	1,059	**13,274**
7	1,364	2,071	1,693	1,719	1,164	**8,011**
8	3,262	7,247	4,261	3,015	2,270	**20,055**
9	5,857	13,825	7,637	5,117	3,112	**35,548**
Totals	**21,587**	**48,624**	**27,910**	**20,200**	**12,665**	**130,986**

5. Costumers with lowest income which is “less than 1499 DM” is the second highest buyer of most expensive coffee packet.

Price_per_Packet \ Income	1	2	3	4	Totals
1	7,509	8,293	9,130	4,705	**29,637**
2	16,914	17,736	21,298	11,426	**67,374**
3	9,057	8,514	10,129	6,275	**33,975**
Totals	**33,480**	**34,543**	**40,557**	**22,406**	**130,986**

6. Costumers with income within the range of “2500 to 3499 DM” consumes coffee the most which is 40557 and the probability is 30.96%.

Price_per_Packet \ Income	1	2	3	4	Totals
1	7,509	8,293	9,130	4,705	**29,637**
2	16,914	17,736	21,298	11,426	**67,374**
3	9,057	8,514	10,129	6,275	**33,975**
Totals	**33,480**	**34,543**	**40,557**	**22,406**	**130,986**

7. If compared between Socio economic level and Price per packet then “upper class” is the second highest buyer of packets whose price range is “6,50 DM to 8,50 DM” which is 12311.

SEC \ Price_per_Packet	1	2	3	Totals
1	5,080	12,311	6,812	**24,203**
2	4,564	11,397	6,341	**22,302**
3	11,519	25,858	11,675	**49,052**
4	5,935	13,109	6,786	**25,830**
5	2,539	4,699	2,361	**9,599**
Totals	**29,637**	**67,374**	**33,975**	**130,986**

8. When comparing brand and loyalty it shows that, Brand 2 "Jacobs other" and 5 "Eduscho Gala" have more "not loyal" observations than "loyal" observations which shows more negative approach rather than positive.

Loyalty \ Brand	1	2	3	4	5	6	7	8	9	Totals
1	8,308	3,500	8,608	5,263	2,816	7,551	4,313	12,881	21,817	**75,057**
2	7,354	5,372	5,224	4,107	3,546	5,723	3,698	7,174	13,731	**55,929**
Totals	**15,662**	**8,872**	**13,832**	**9,370**	**6,362**	**13,274**	**8,011**	**20,055**	**35,548**	**130,986**

9. Age group 4 which is "50 to 59 years" is the highest buyer (13082) of the coffee brand which is "Andere Kaffeemarken" and probability is 35.57%.

Age \ Brand	1	2	3	4	5	6	7	8	9	Totals
1	276	283	359	284	114	213	179	401	566	**2,675**
2	1,892	1,214	1,898	1,430	832	1,513	1,339	2,733	4,093	**16,944**
3	2,405	1,897	2,880	1,915	1,518	2,827	1,800	4,299	7,208	**26,749**
4	5,439	2,787	4,708	3,669	2,143	4,656	2,727	7,390	13,082	**46,601**
5	5,650	2,691	3,987	2,072	1,755	4,065	1,966	5,232	10,599	**38,017**
Totals	**15,662**	**8,872**	**13,832**	**9,370**	**6,362**	**13,274**	**8,011**	**20,055**	**35,548**	**130,986**

10. Costumers income range between “2500 DM - 3499 DM” is the highest (11564) for brand 9 which is "Andere Kaffeemarken" but the income of people which is lowest within the range "less than 1499 DM" (8963) is the second buyer of the same brand.

Brand	Income	1	2	3	4	Totals
1		4,699	4,053	4,605	2,305	**15,662**
2		2,207	2,350	2,759	1,556	**8,872**
3		2,750	4,424	4,699	1,959	**13,832**
4		2,249	2,710	3,069	1,342	**9,370**
5		1,654	1,490	1,631	1,587	**6,362**
6		3,162	3,048	3,766	3,298	**13,274**
7		2,470	1,846	2,611	1,084	**8,011**
8		5,326	5,679	5,853	3,197	**20,055**
9		8,963	8,943	11,564	6,078	**35,548**
	Totals	**33,480**	**34,543**	**40,557**	**22,406**	**130,986**

11. The most famous brand is “Andere Kaffeemarken” (35548) followed by “Tchibo other” (20055) and “Jacobs Krönung” (15662) respectively. Probability that people are consuming these three brands is (0.528797) 52.87 % which is almost half of total brands consumed.

Brand	Totals
1	**15,662**
2	**8,872**
3	**13,832**
4	**9,370**
5	**6,362**
6	**13,274**
7	**8,011**
8	**20,055**
9	**35,548**
Totals	**130,986**

12. Comparing between no of packet and price per packet price range within “6,50 – 8,50 DM” is highest for 1 packet which is (51892). Probability of customers buying “one packet ” is 80.17% (105022) and probability of customers buying highest price packet which lies within range “more than 8,50 DM” is 25.93 which is second highest (29888) shows good symbol for business because 26% of the people are buying the most expensive packet.

Price_per_Packet / No_of_Packet	1	2	3	Totals
1	23,242	51,892	29,888	**105,022**
2	4,504	11,410	3,375	**19,289**
3	1,891	4,072	712	**6,675**
Totals	**29,637**	**67,374**	**33,975**	**130,986**

13. Data of IDNO seems irrelevant to me till now however I have done some calculations to find some pattern out of it.

```
> range(datagiven$IDNo)
[1]    1 7251
> names(which.max(table(datagiven$IDNo)))
[1] "560"
> sum(datagiven$IDNo == 560)
[1] 251
```

The above data shows that the range of IDNO lies between 1 to 7251 and IDNO = “560” is most repeated which is 251 times.