1. How can data science help businesses?

Data Science has these benefits for businesses:

Empowering management and officers to make better decision

Directing actions based on trends—which in turn help to define goals

Challenging the staff to adopt best practices and focus on issues that matter

Identifying opportunities

Decision making with quantifiable, data-driven evidence

Testing these decisions

Identification and refining of target audiences

Recruiting the right talent for the organization

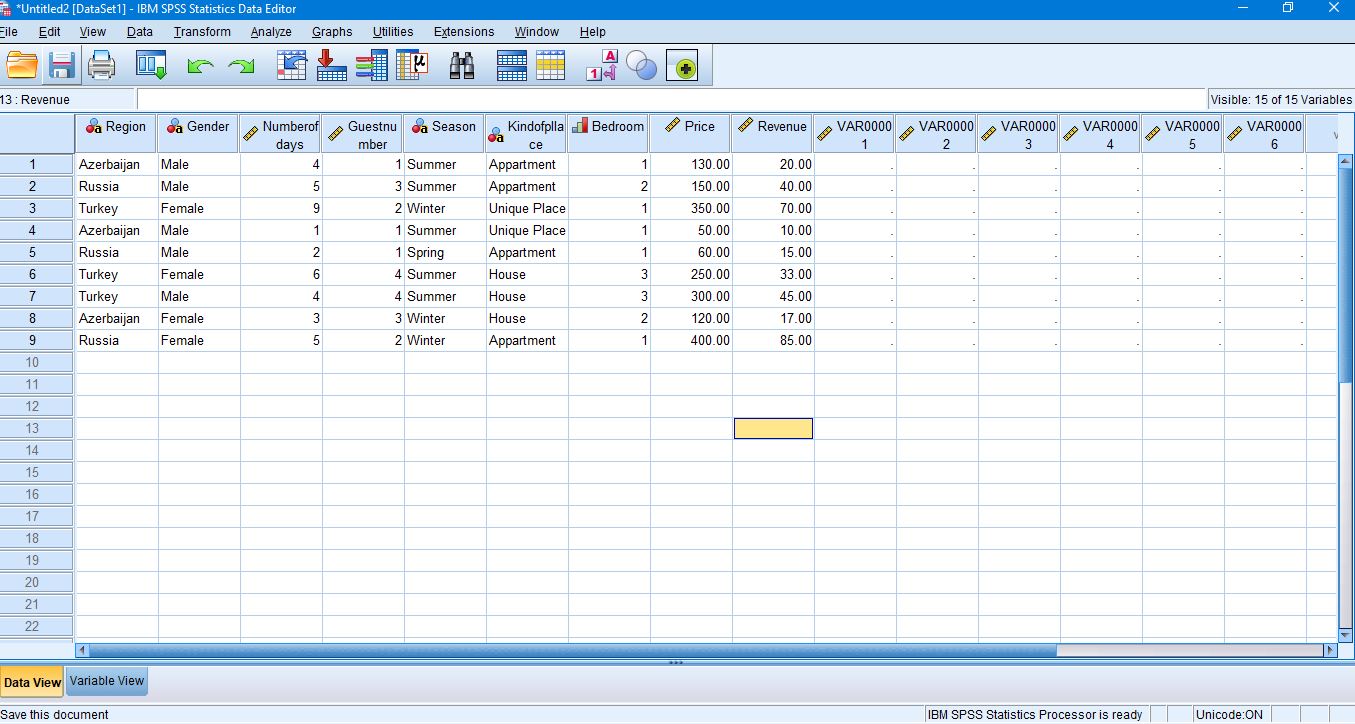
For example, in 2014, Airbnb found out that users from certain countries would click the neighborhood link, browse the page and photos and not make any booking.

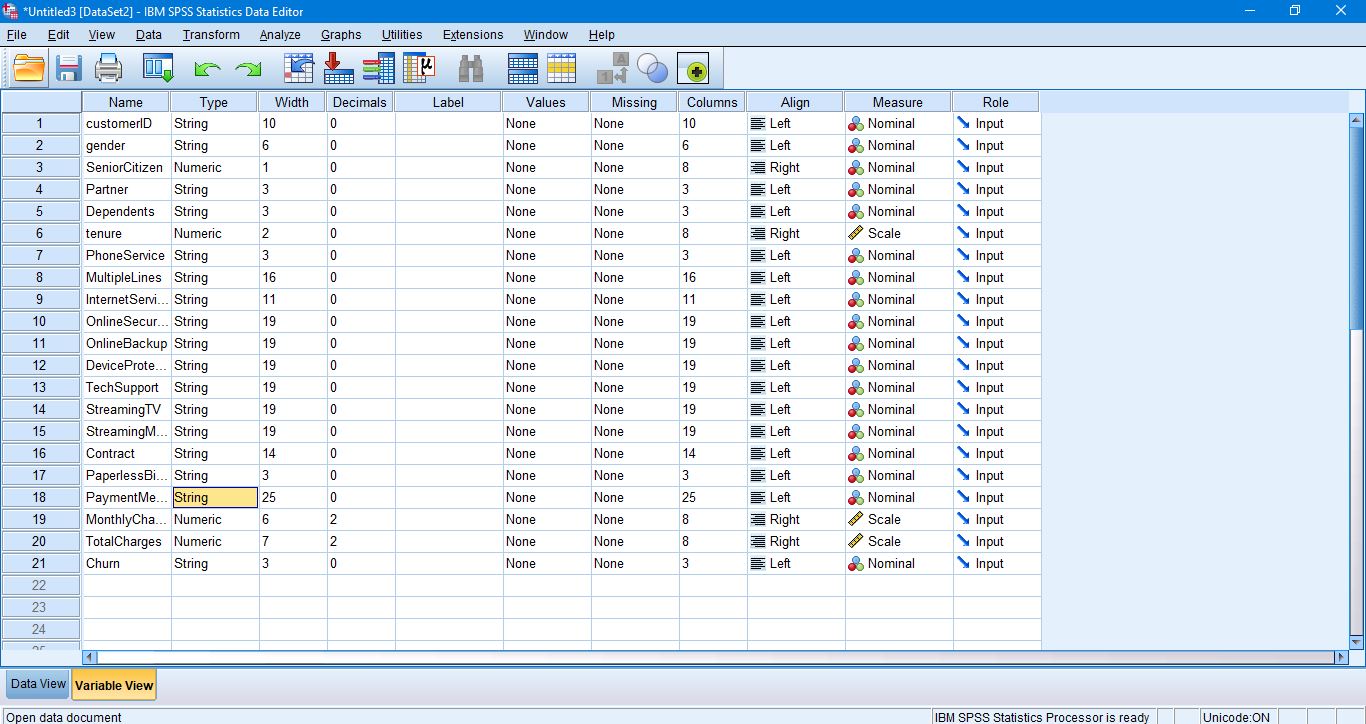
In order to mitigate this issue, Airbnb released a different version for the users from those countries and replaced neighborhood links with the top travel destinations. This saw a 10% improvement in the lift rate for those users.

2. For your example list possible important variables to get meaningful insight for the company. You may conduct internet research to find out for popular data science use cases and commonly used variables for them.

Users’ (subscriber and host) region (categorical variable), trip dates (interval), guest number (quantitative), prices (quantitative), season (categorical variable) and kind of places (nominal) is important variables for my example.

3. Indicate numerical, categorical, nominal and ordinal variables in your list.

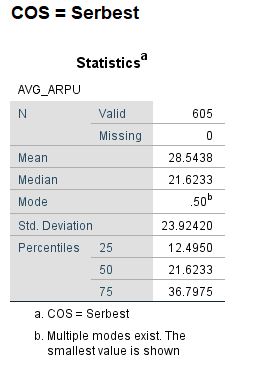
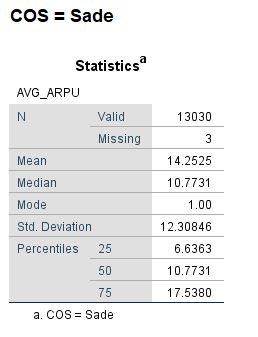
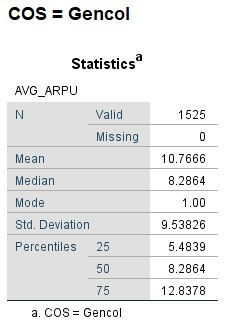




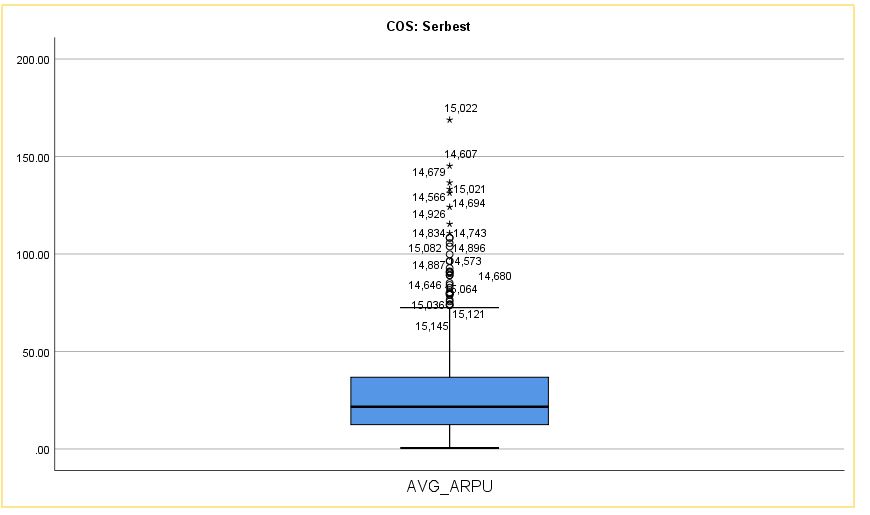
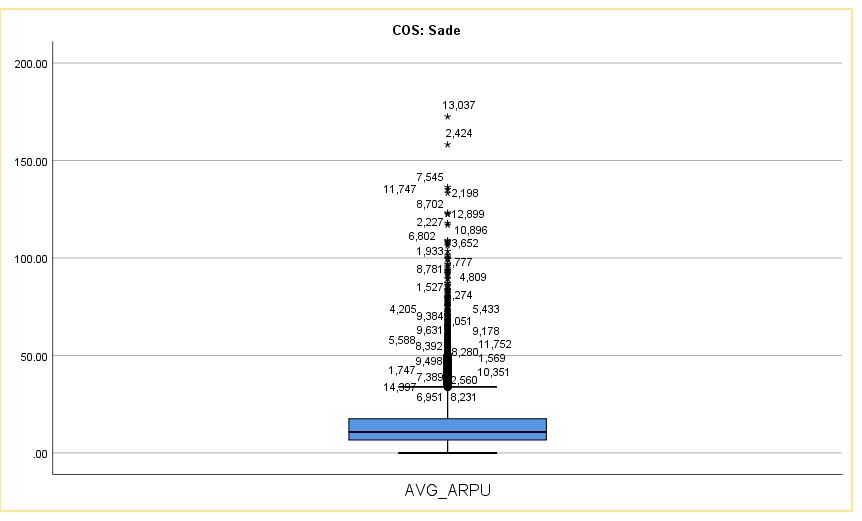
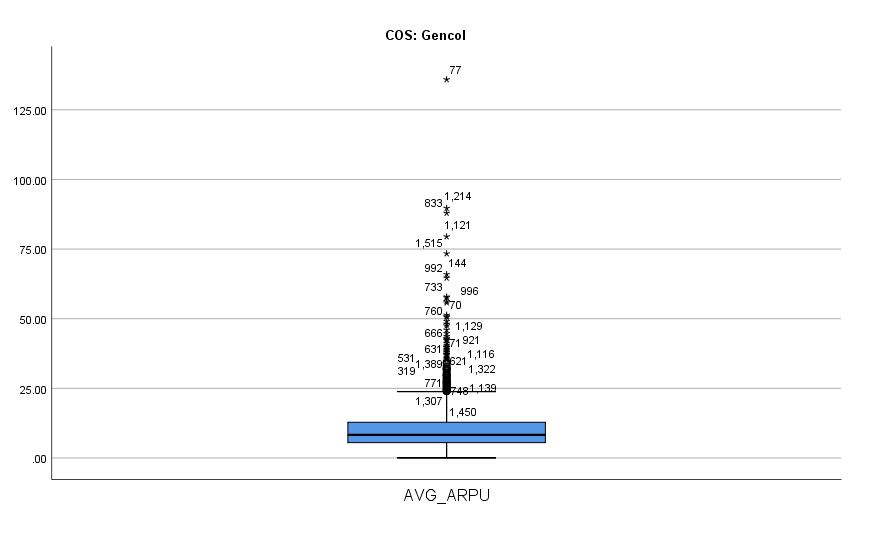
4. Explain importance of Statistics to solve your business problem. Use fundamental statistical terms (sample, population, inference, significance and sampling) in your explanation.

Population is the entire set of possible cases. Sample is a subset of the population from which data are collected. In this case people from Azerbaijan, Turkey and Russia are samples. We made a statistical analyze for these regions and statistical inference is to estimate this sample to sample variation or uncertainty. Understanding how much our results may differ if we did the study again, or how uncertain our findings are, allows us to take this uncertainty into account when drawing conclusions. Statistical significance is a determination made by an analyst that the results in the data are not explainable by chance alone. Statistical hypothesis testing is the method by which the analyst makes this determination. This test provides a p-value, which is the probability of observing results as extreme as those in the data, assuming the results are truly due to chance alone. A p-value of 5% or lower is often considered to be statistically significant.

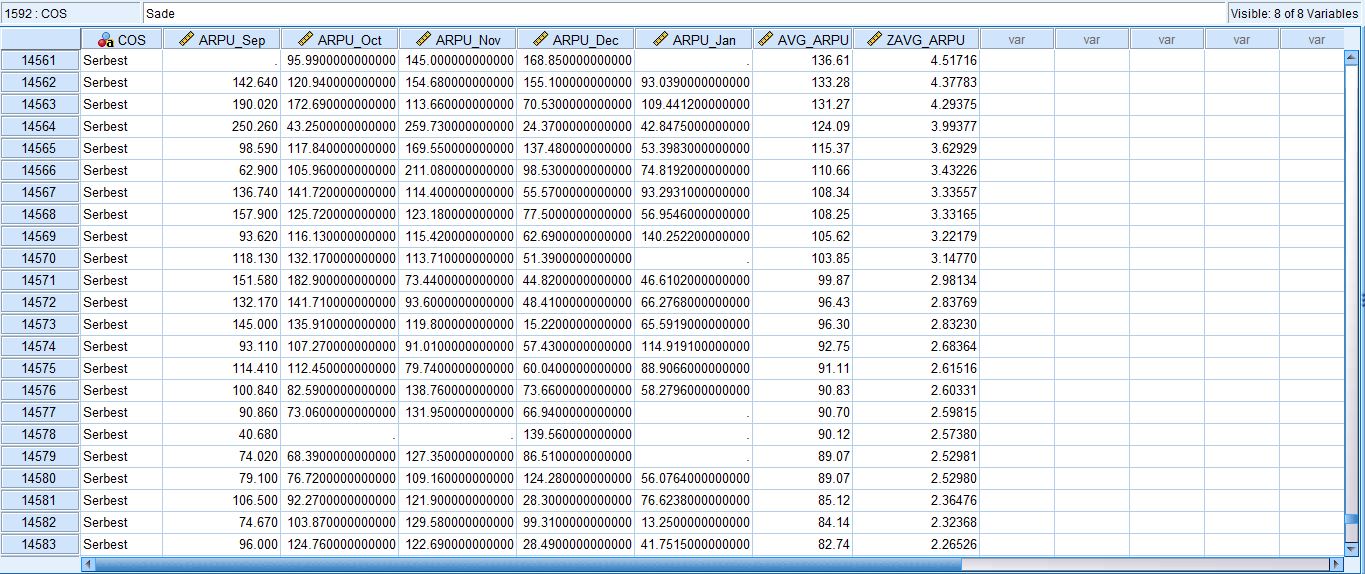
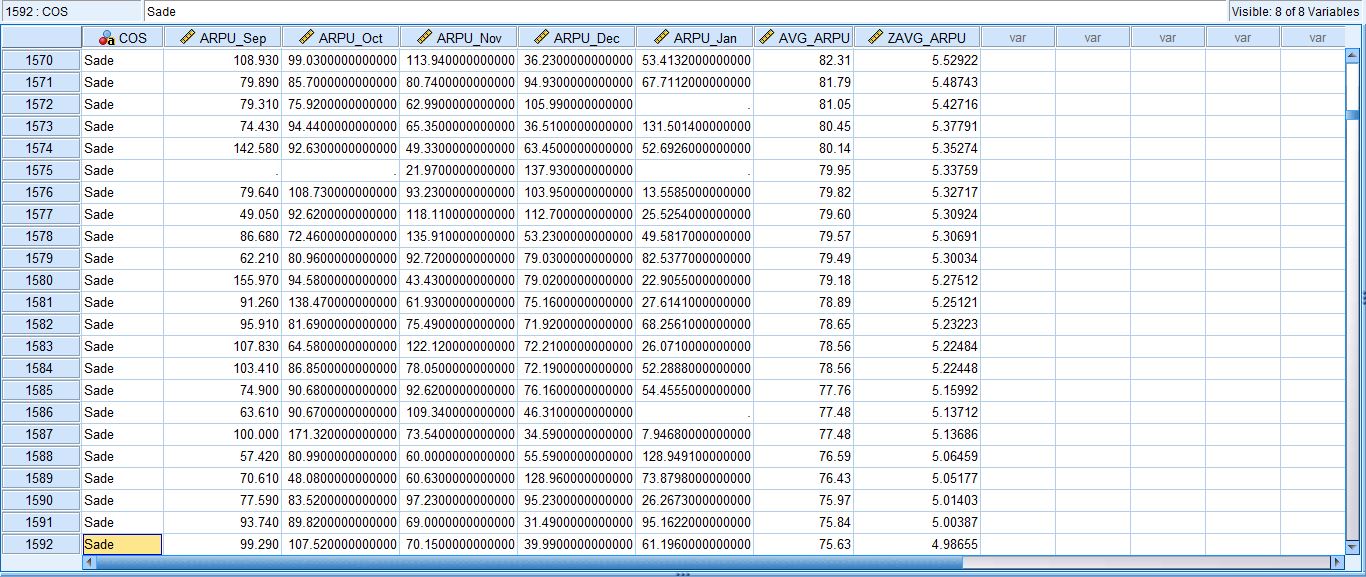
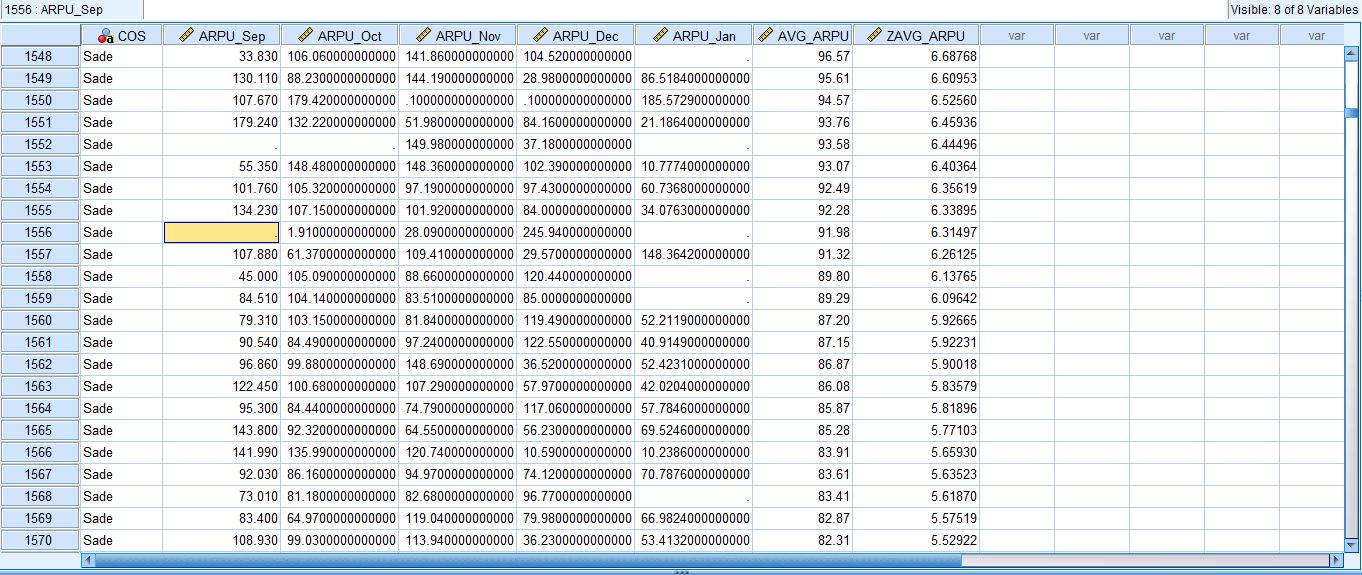
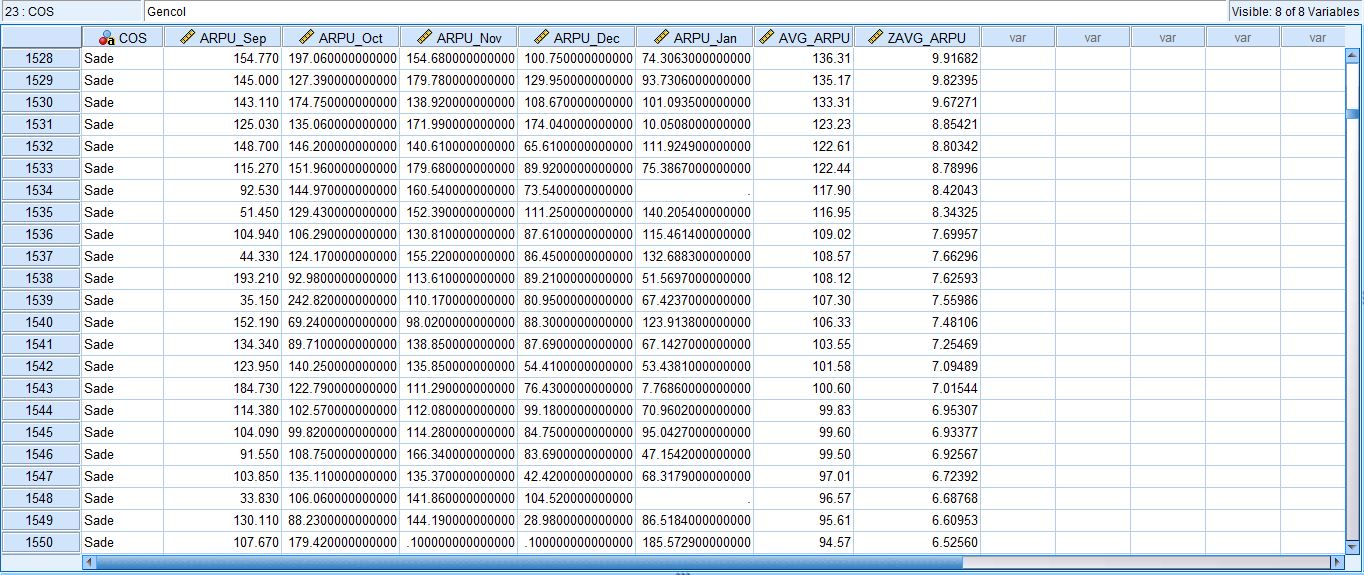
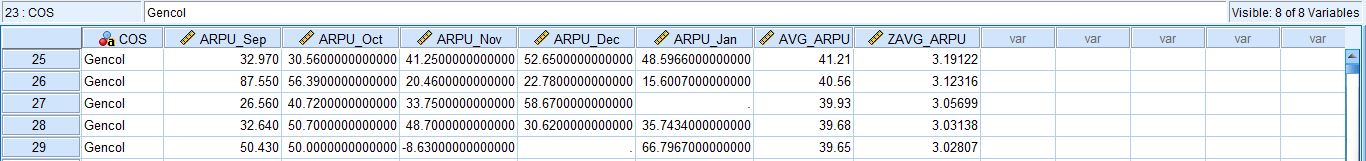
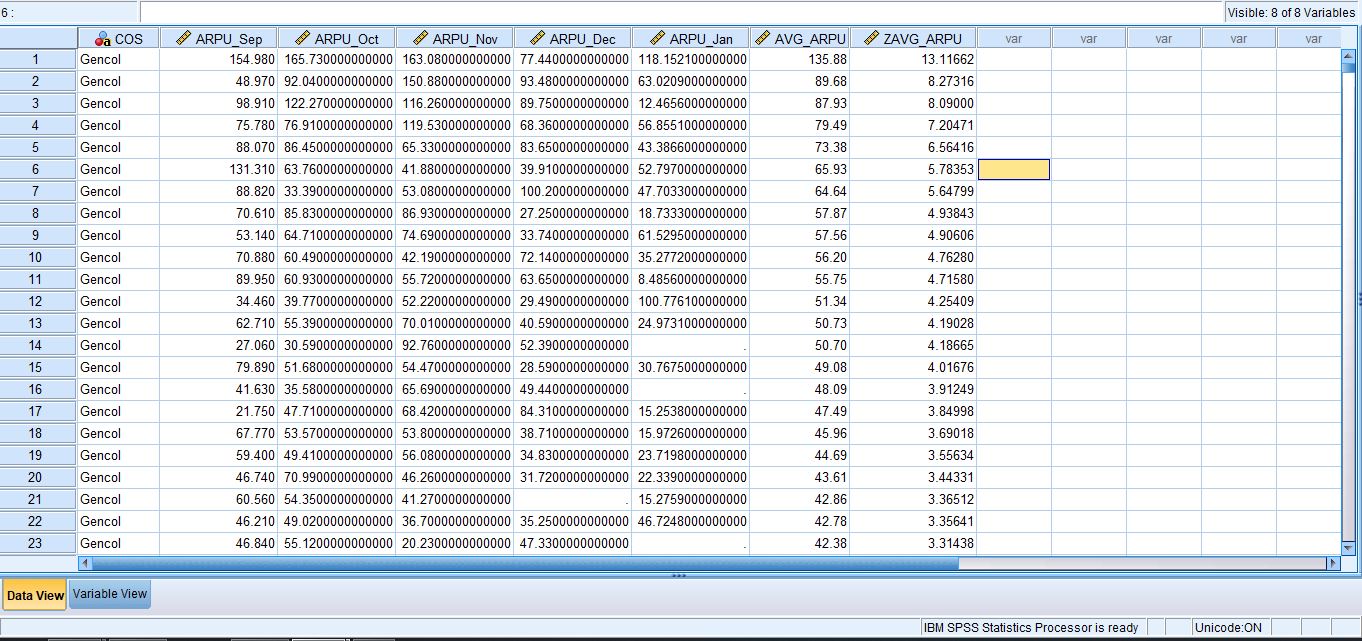
5. Describe average revenue per user for each COS (class of service) in telecommunication data. Use box plot for your analysis and include mean, median, mode, quartiles, percentiles and standard deviation in your report. Explain which COSes are performing well and poor.



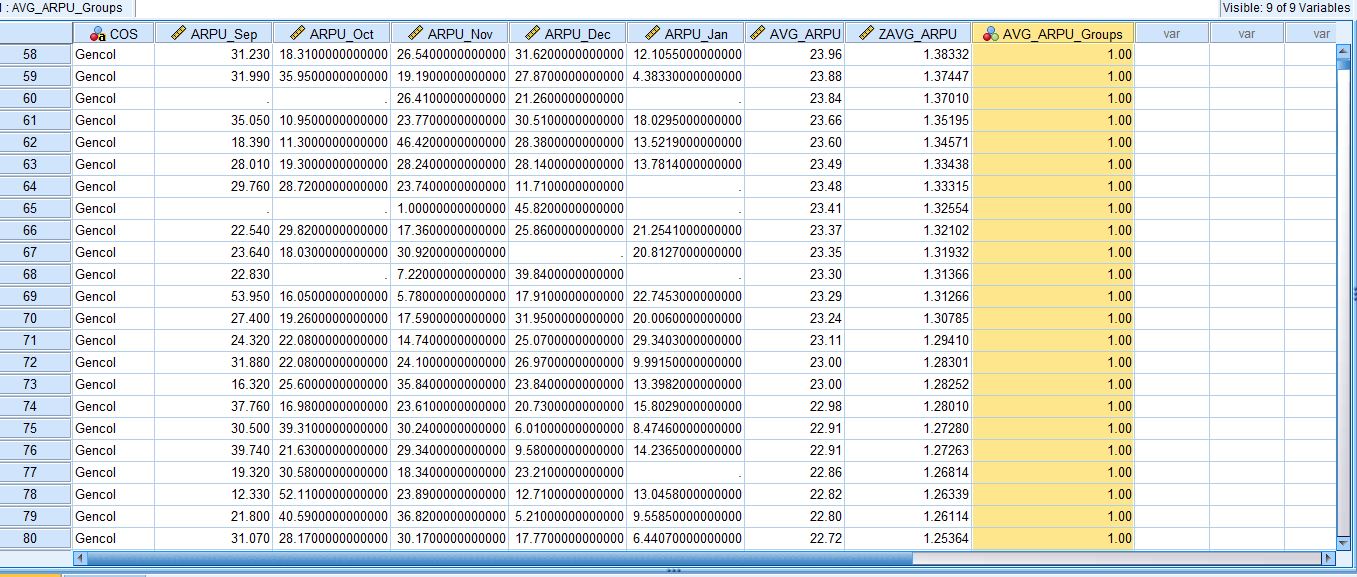
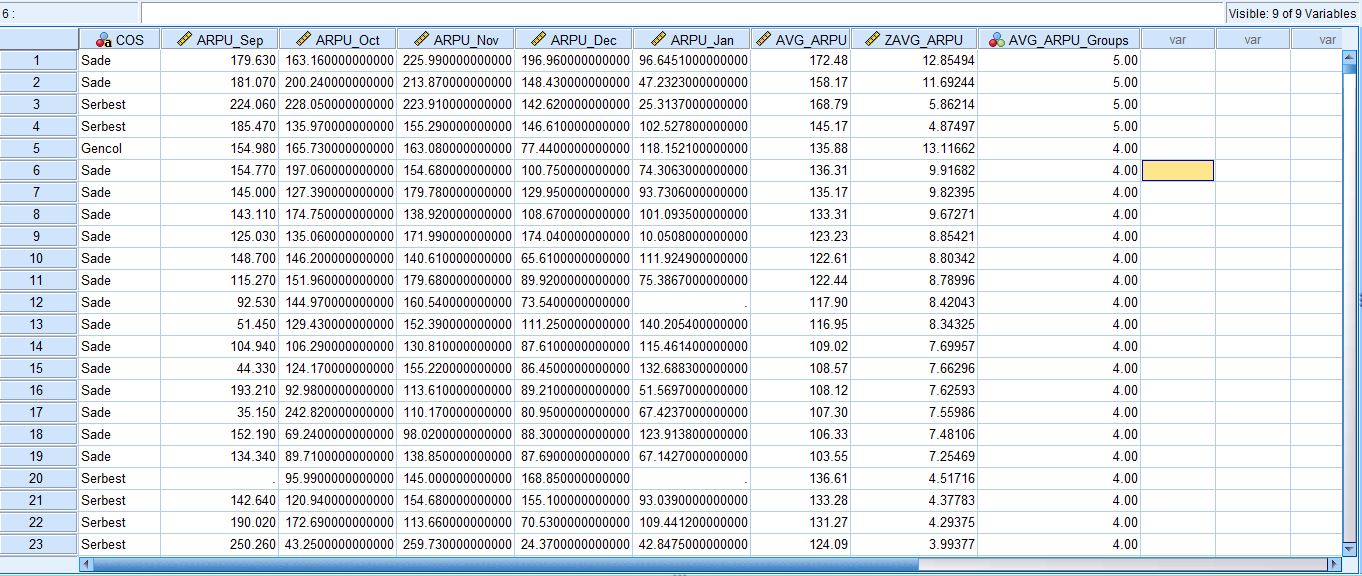
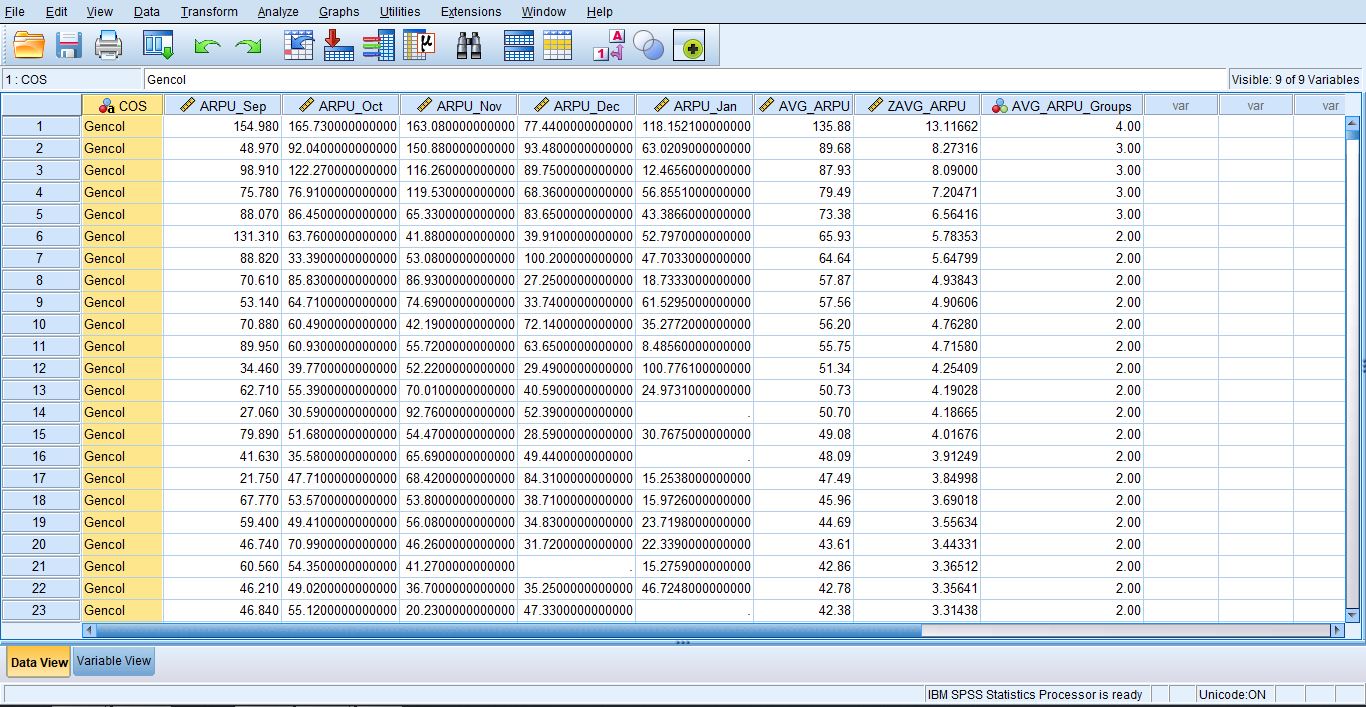
Gencol is performing well, because its Std. Deviation is low. Serbest is performing poor, because its Std. Deviation is high.



6. Show outliers for each COS.



7. For each COS recode average revenue to 5 different subcategory. Data range should be split equally according to distribution of data.\*



8. Open Telecommunication\_Case\_1.sav in SPSS Modeler and create a table that lists all values in your data.

9. Provide a comprehensive first look at the data using Data Audit node.

