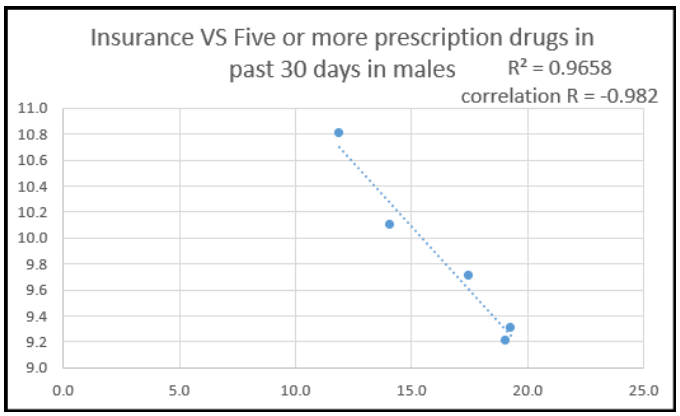
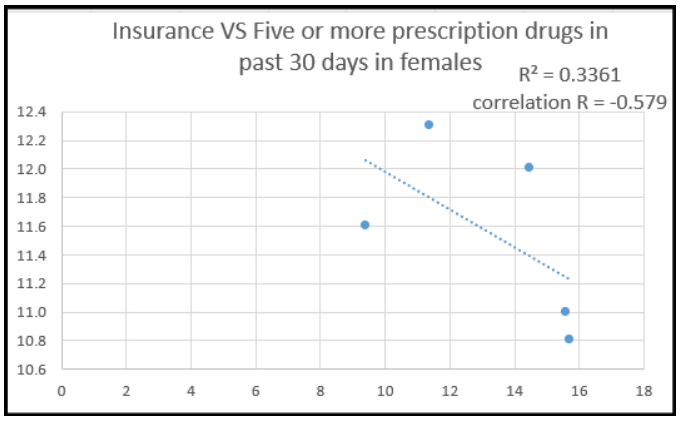
Data Integration and Data Mining (HOA#6)

We have taken the data of prescription of drugs of (at least one in last 30 days, five and more in last 30 days) and the insurance on the basis of sex from year 2007 – 2018. Data was integrated together by taking the average over 3 years in the insurance data because it was given yearly wise but the prescription data was in the range of years. We converted the yearly data to range of years by taking the average over those years.

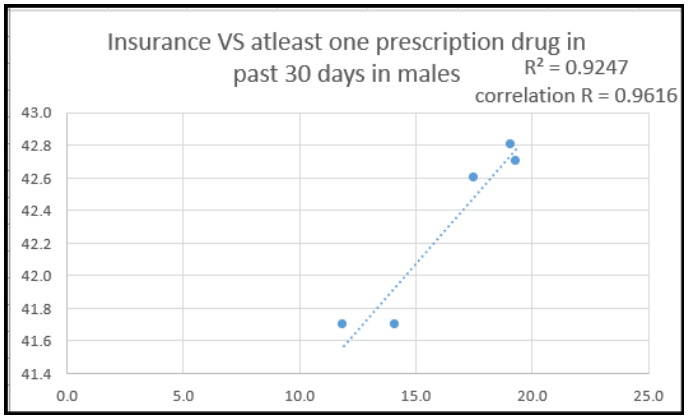
1. **Is there a correlation between drug use and insurance coverage?**
   Yes, there is a correlation between the drug use and insurance coverage. Here is scatter plot and the trendline representing that



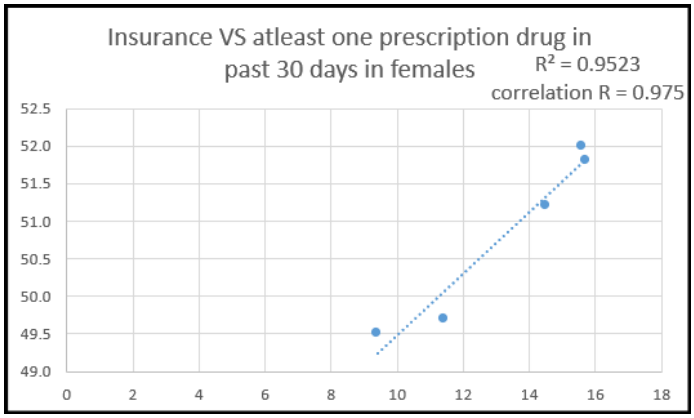
Negatively related



Negatively related



Positively related



Positively Related

2. **Is there a trend, do you see clusters?**
   Yes, there is a trend as the health insurance increases in males and females at least one prescription drug also increases in them especially more in female as their correlation value is higher than males but the five or more prescription decreases with the increase in health insurance. It decreases more in case of males.

3. **This exercise is an example of which generation of BI?**
   Since this exercise include the simple visualization like scatterplots it is the first generation of the BI.

4. **What is the advantage if users can do something like this without support from IT (= self-service BI)?**
   Reduced IT Burden:
   IT teams are freed from mundane data analysis tasks, allowing them to focus on more complex technical issues and strategic initiatives.
   Cost Savings:
   Reduced reliance on IT support for data analysis can potentially lead to cost savings in the long run.
   Empowerment and Ownership:
   Business users feel empowered to explore data relevant to their roles, fostering a data-driven culture and promoting ownership of data analysis.

5. **What about the "single source of truth" in the context of this type of BI? (Data Governance)**
   Approaches to a Single Source of Truth:
   Centralized Data Warehouses: The concept of data warehouses emerged, aiming to provide a central repository for storing integrated data from various sources.
   Standardization Efforts: Organizations might define data standards and dictionaries to ensure consistency in how data was captured and stored across different systems.
   Data Cleansing and Validation Processes: Efforts were made to clean and validate data before loading it into the data warehouse to improve data quality.