KHUSHI PATEL

MSBA 305

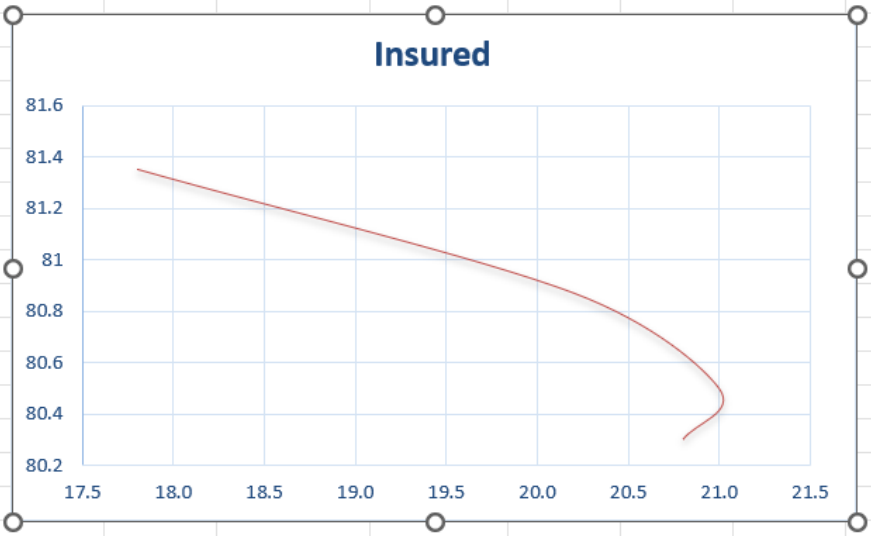
Data Integration and Data Mining HOA6

For this assignment, the data adopted covers a period from 1999 to 2008. This gives a period of the entire decade, hence enough view on the trend and pattern of prescription drug usage without compulsory health insurance coverage involved.

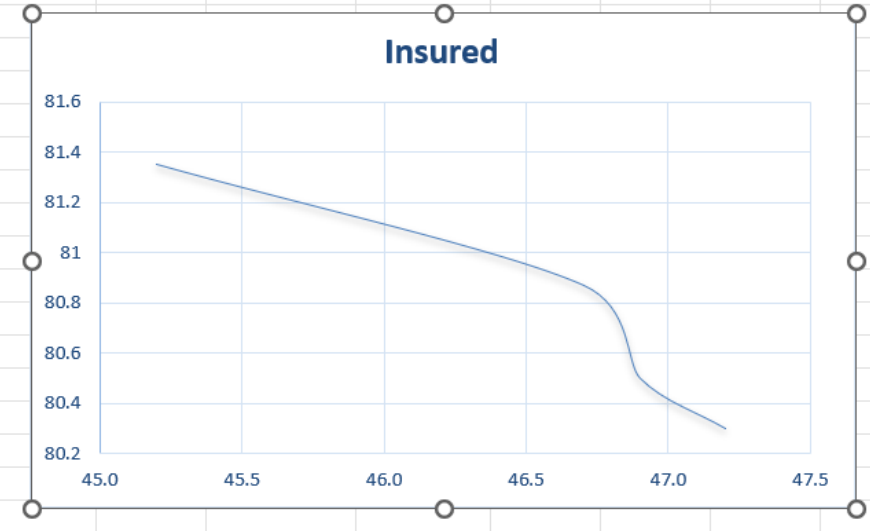
Year	1+ Drug Use	3+ Drug Use	5+Drug use	Insured
1999-2002	45.2	17.8	7.5	81.85
2001-2004	46.7	20.2	9.2	80.87
2003-2006	46.9	21.0	10.0	80.5
2005-2008	47.2	20.8	10.2	80.3

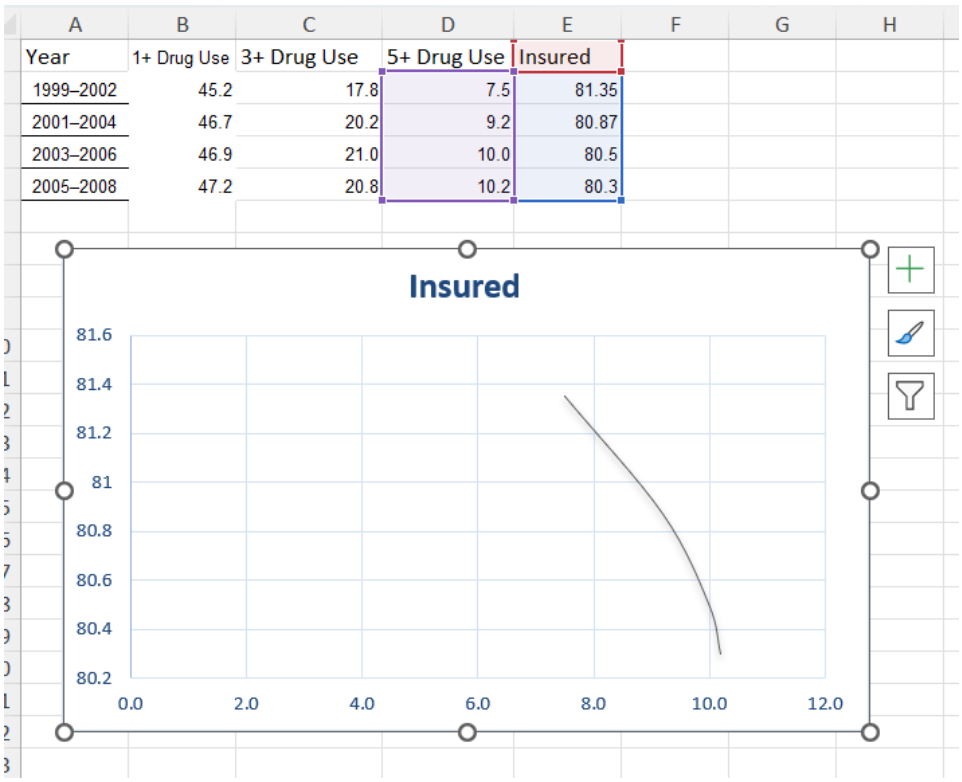
To correlate the use of prescription drugs with health insurance coverage, three images are included in this analysis. The first image is about members who used at least one prescription drug (1+ drug use), illustrating how insurance coverage provides greater access to medications. The second image is about members who took three or more prescription drugs (3+ drug use), indicating a stronger connection between insurance and increased medication use. The third image is of people who used five or more prescription drugs (5+ drug use), showing the strongest correlation, where better insurance coverage leads to greater use of multiple medications. Together, these images demonstrate how insurance coverage favors the increased use of prescription drugs.

Year	1+ Drug Use	3+ Drug Use	5+ Drug Use	Insured
1999–2002	45.2	17.8	7.5	81.35
2001–2004	46.7	20.2	9.2	80.87
2003–2006	46.9	21.0	10.0	80.5
2005–2008	47.2	20.8	10.2	80.3



Year	1+ Drug Use	3+ Drug Use	5+ Drug Use	Insured
1999–2002	45.2	17.8	7.5	81.35
2001–2004	46.7	20.2	9.2	80.87
2003–2006	46.9	21.0	10.0	80.5
2005–2008	47.2	20.8	10.2	80.3





1. **Is there a correlation between drug use and insurance coverage?**

   The correlation is negative, that is, if prescription drug use increases or if 3+ or 5+ drugs are used more commonly, the percentage of insured persons decreases, implying an inverse relationship.

2. **Is there a trend, and do you see clusters?**

   Yes, there is indeed a specific trend. The data present rather show that when the coverage is high, people make such use of prescription drugs. There are various clusters, low and high: high group uses the drugs with high insurance coverage, while there are low clusters with low coverage.

3. **This exercise is an example of which generation of BI?**

   In the second generation of BI, users have access to self-service tools such as Excel or Tableau, through which they may analyze and visualize data without needing IT support

every time. With respect to this new generation of BI, users, without IT support, are now using available self-service tools like Excel or Tableau to analyze and visualize data. This embodiment is indicative of Second Generation BI. This generation is characterized by self-service tools, like Excel or Tableau, which allow users to analyze and visualize data without having to call or consult frequently with the IT department.

4. **What is the advantage if users can do something like this without support from IT (Self-Service BI)?**

   Speed is the obvious advantage. Users can quickly combine data, find trends, and generate answers on their own. They will not have to rely on IT for processing the data, thereby saving their own time and enabling faster decision-making.

5. **What about the "single source of truth" in the context of this type of BI? (Data Governance)**

   So, the main challenge occurs when everybody works with the data separately; that is why we need systems such as a shared database to ensure that all are operating on the same trustworthy version of the data. This way, consistency and reliability of the insights are assured.