**DLK GROUP**

**Dataset History**

The dataset represents ten years (1999-2008) of clinical care at 130 US hospitals and integrated delivery networks. Each row concerns hospital records of patients diagnosed with diabetes, who underwent laboratory, medications, and stayed up to 14 days. The goal is to determine the early readmission of the patient within 30 days of discharge. The problem is important for the following reasons. Despite high-quality evidence showing improved clinical outcomes for diabetic patients who receive various preventive and therapeutic interventions, many patients do not receive them. This can be partially attributed to arbitrary diabetes management in hospital environments, which fail to attend to glycemic control. Failure to provide proper diabetes care not only increases the managing costs for the hospitals (as the patients are readmitted) but also impacts the morbidity and mortality of the patients, who may face complications associated with diabetes.

**Dataset Brief**The dataset represents ten years (1999-2008) of clinical care at 130 US hospitals and integrated delivery networks. It includes over 50 features representing patient and hospital outcomes. Information was extracted from the database for encounters that satisfied the following criteria. (1) It is an inpatient encounter (a hospital admission). (2) It is a diabetic encounter, that is, one during which any kind of diabetes was entered into the system as a diagnosis. (3) The length of stay was at least 1 day and at most 14 days. (4) Laboratory tests were performed during the encounter. (5) Medications were administered during the encounter. The data contains such attributes as patient number, race, gender, age, admission type, time in hospital, medical specialty of admitting physician, number of lab tests performed, HbA1c test result, diagnosis, number of medications, diabetic medications, number of outpatient, inpatient, and emergency visits in the year before the hospitalization, etc.

**TASK:**

**Dataset**: Diabetes 130-US Hospitals for Years 1999–2008

**Dataset link:** <https://drive.google.com/drive/folders/1z72doTq-V72RfIIXjeD9CX2vxVfAiS9M>

**Objective**:

To analyse hospital readmissions of diabetic patients and identify patterns, trends, and risk factors that influence patient outcomes and hospital efficiency.

**Overview:**

You are required to analyse the dataset and generate actionable insights into patient readmissions, treatment patterns, and hospital performance. The goal is to demonstrate your ability to explore healthcare data and deliver both technical findings and strategic recommendations.

**Expected Insights:**

***1. Readmission & Risk Factors***

* Which variables are most predictive of early readmission (within 30 days)?
* What percentage of diabetic patients are readmitted within 30 days?
* How does the number of inpatient/emergency/outpatient visits affect readmission risk?
* Do patients with abnormal A1C or max glucose serum levels have higher readmission rates?
* What percentage of patients were readmitted within 30 days?
* What features are the strongest predictors of early readmission?
* How does the number of prior inpatient visits correlate with readmission?
* Is there a threshold of emergency visits that significantly increases readmission likelihood?
* Are more outpatient visits associated with better or worse outcomes?
* How do A1C test results relate to early readmission?
* Are abnormal max glucose serum levels predictive of readmission?
* Do patients with more medications tend to be readmitted more often?
* Is the combination of abnormal lab values and emergency visits a strong indicator of readmission?
* Can a logistic regression model accurately classify early readmissions?
* How well do tree-based models (e.g., Random Forest) perform in predicting readmission?
* Can we build a readmission risk score to categorize patients (low, medium, high risk)?

***2. Demographic Disparities***

* Are certain races or age groups more likely to be readmitted?
* Do gender differences exist in length of stay or frequency of readmission?
* Are there disparities in the prescription of diabetes medications among different demographic groups?
* Which racial groups have the highest 30-day readmission rates?
* Are older age groups more prone to readmission?
* How do gender-based readmission rates compare?
* Are length-of-stay differences statistically significant across genders?
* Do specific race and gender combinations experience higher-than-average readmission?
* Are minorities under-prescribed A1C tests or diabetes medications?
* Do men and women differ in the number of lab procedures received?
* Are treatment changes more likely for certain demographic groups?
* Are younger patients (e.g., under 30) admitted with more severe conditions?
* Does age influence the likelihood of being discharged home vs. a facility?
* Do readmission rates vary significantly across age brackets (e.g., 50–60 vs. 70–80)?
* Are older patients more likely to receive insulin or combination drugs?

***3. Treatment & Medication Insights***

* Which diabetes medications are associated with the lowest readmission rates?
* Does changing medication during a hospital stay lead to better patient outcomes?
* How does insulin usage correlate with hospital stay or readmission?
* What is the impact of combination drugs (e.g., glyburide-metformin) on patient outcomes?
* Which individual medications are associated with the lowest readmission rates?
* Do combination therapies reduce readmission more effectively than single drugs?
* Are insulin users more likely to be readmitted than non-insulin users?
* Does metformin use correlate with shorter hospital stays?
* Are there medications frequently changed during the stay that reduce risk?
* Do patients who received "No" for diabetes medication have worse outcomes?
* Which medications are more often associated with abnormal A1C values?
* What is the average number of medications for readmitted vs. non-readmitted patients?
* Are certain medications only prescribed to certain demographics?
* What medication patterns are most common in non-readmitted patients?
* Is there an interaction between medication type and number of diagnoses?
* Do patients who experienced a medication change have better A1C improvements?

***4. Hospital Performance & Process Analysis***

* Which admission types (e.g., emergency vs. elective) are associated with the highest readmission rates?
* How do discharge dispositions affect the likelihood of readmission?
* Which medical specialties are most effective at preventing readmissions?
* Is there a correlation between the number of lab procedures and better health outcomes?
* Which admission types have the highest readmission rates?
* Are elective admissions less likely to be readmitted than emergency ones?
* How does admission source (e.g., referral vs. emergency room) affect readmission?
* What discharge disposition types correlate with highest readmission risk?
* Do patients discharged to skilled nursing facilities have longer stays?
* Which medical specialties are associated with the lowest readmissions?
* Are certain specialties more aggressive in ordering lab tests?
* Does medical specialty influence length of stay or medication usage?
* Are patients in general practice less likely to receive A1C testing?
* Which discharge categories (e.g., expired, home, transfer) relate to patient mortality?
* Does frequency of lab procedures reflect better outcomes?
* Can we rank hospitals or departments based on performance using readmission and length of stay?

***5. Length of Stay Analysis***

* What are the top predictors of longer hospital stays for diabetic patients?
* Is there a relationship between length of stay and readmission?
* Do certain diagnoses (diag\_1/2/3) lead to significantly longer hospitalizations?
* What is the distribution of hospital stays for diabetic patients?
* What variables most strongly influence the length of hospital stay?
* How does time in hospital correlate with readmission risk?
* Do more lab procedures increase or decrease length of stay?
* How does length of stay differ across medical specialties?
* Are longer stays associated with better A1C or glucose control?
* Is there a point beyond which a longer stay no longer reduces readmission?
* What diagnoses are most associated with prolonged hospitalization?
* Do certain medications correlate with shorter stays?
* Are readmitted patients typically those who stayed shorter or longer initially?
* What is the average length of stay by admission type?
* Does a longer stay predict higher number of procedures or tests?

***6. Diagnosis & Comorbidity Insights***

* What are the most common comorbid diagnoses among diabetic patients?
* How does the number of diagnoses relate to readmission or mortality risk?
* Are there certain diagnosis patterns that indicate high-risk diabetic patients?
* What are the top 10 most frequent primary diagnoses (diag\_1)?
* Which combinations of diag\_1, diag\_2, diag\_3 are most common?
* Which diagnosis combinations are associated with the highest readmission rates?
* Do comorbidities like hypertension or kidney disease influence readmission?
* How does the number of diagnoses relate to time in hospital?
* What are the most common comorbid patterns among patients who died?
* Are certain diagnoses more common in younger vs. older patients?
* How do patients with 9+ diagnoses differ in outcomes from those with <3?
* Which diagnostic categories correlate most with abnormal glucose levels?
* Do patients with heart-related diagnoses stay longer than others?
* Can we predict likely readmission based solely on diagnosis codes?
* Are certain diagnosis combinations tied to higher medication counts?

***7. Utilization & Resource Efficiency***

* How does the number of medications or lab procedures relate to outcomes?
* Are more procedures always linked to better results, or is there a point of diminishing returns?
* What is the average resource utilization (labs, procedures) for patients who are not readmitted?
* How does the number of lab procedures relate to readmission risk?
* Do patients with more medications always fare better?
* Is there a diminishing return effect in number of procedures and outcomes?
* What is the average number of lab procedures for non-readmitted patients?
* How does resource usage vary by age or gender?
* Are high-utilization patients receiving better care, or just more interventions?
* Do more procedures correlate with longer stays or fewer readmissions?
* Is there a consistent pattern in lab tests for patients with normal A1C results?
* Do patients with the same diagnosis receive varying levels of care?
* Are there outliers with very high resource use and poor outcomes?
* Is there a “sweet spot” of lab tests/medications that minimizes readmission?
* Can we cluster patients based on utilization patterns and outcome?

***8. Temporal Patterns***

* How have readmission rates changed over the 1999–2008 period?
* Has the use of specific medications (e.g., metformin) increased or decreased over time?
* Are there improvements in treatment protocols over the 10-year span?
* How did readmission rates trend from 1999 to 2008?
* Was there an increase in average length of stay over time?
* Did metformin or insulin usage increase or decrease over the years?
* Are newer drugs (e.g., rosiglitazone) more common in later years?
* Were there specific years with spikes in lab test frequency?
* Has the number of diagnoses per patient increased over the years?
* Did outcomes improve as treatment protocols evolved over time?
* Were treatment changes more frequent in early or later years?
* Were earlier years associated with higher mortality or longer stays?
* How did the use of combination therapies change over time?
* Was there a shift in dominant diagnosis codes across years?
* Are more recent records associated with lower readmission and better A1C control?
* **Outcome Required on 17-05-2025 (Saturday)**