חלק א:

1. בחרנו בסכימת Snowflake - בגלל שיש לנו Fact Table יחיד וישנן עוד טבלאות שלא מתחברות ל-Fact Table באופן ישיר.
2. טבלת ה-Fact שלנו היא טבלת ה-Events שמתארת את האירועים השונים שמתרחשים במשחקים השונים, בכל אירוע יש מסווג משלו ומסווג של המשחק (מטבלה Matches).
ישנה טבלה Attempts שכאשר האירוע שקורה הוא ניסיון בעיטה לשער אז המסווג שלו יופיע גם כן בטבלת Attemps עם מידע נוסף בהקשר לניסיון הבעיטה לשער.
הטבלה Events תקושר לטבלה בשם Matches. ל-Match יהיה מסווג משלו, מסווג קבוצה מארחת, מסווג קבוצה מתארחת והעונה בה התרחש המשחק.
הטבלה Matches תקושר ל-2 טבלאות - Seasons ו-Teams. כאשר בקבוצה Teams יהיה תיאור של הקבוצות השונות ו-Season יהיה רשימה של העונות השונות.
טבלה Teams תקושר לטבלה Leagues. כאשר בטבלה Teams יהיה מסווג קבוצה ומסווג ליגה.
טבלה Leagues תתאר את הליגות השונות שבהן התרחשו המשחקים.
3. .
4. בעקבות הוספת הטבלה Attempts, פיצלנו את המידע הנוסף שיש לנו על אירועים בהם היה ניסיון בעיטה לשער, כאשר בטבלה המקורית במידה והאירוע לא היה בעיטה לשער אז בעמודות הקשורות במידע של הניסיון בעיטה לשער הופיעו כמויות גדולות של NAs (כ-70%). כך שבמקרה ואנו רוצים לחפש את כל האירועים שהסתיימו בגול אז אנו צריכים לחפש בטבלה קטנה יותר (Attempts) ובכך אנו מקטינים את כמות ה-data.

חלק ב:

1. תהליך הETL:

E- תחילה נחלץ את שתי הטבלאות מה- DATA SOURCE שלנו. הראשונה events.csv המכילה מידע על אירועי המשחק והשניה ginf.csv המכילה מידע על המשחקים עצמם.

T- נוריד את את הנתונים שלא רלוונטים לשאלות העסקיות שהגדרנו (כמו הימורים, שמות שחקנים, זמן במשחק..), נבצע טרנספורמציה לשנות האירועים, נציג במקום תאריך משחק את עונת המשחק. נשנה את שמות הקבוצות הקשורות לאירועים לשמות המקוריים (במקום 0/1).

L- נמלא את השדות הריקים אם יהיו, ונטען את הנתונים לשדות בסכמת ה- DW.

2. עבור טבלת הליגות:
נקח את שמות הליגה והמדינה מטבלת ginf ונעשה distinct כדי לבטל ערכים כפולים.

עבור טבלת הקבוצות:
נקח את שמות הקבוצות, נשייך את הפתח של הליגה לכל קבוצה ונעשה distinct כדי לבטל ערכים כפולים.

עבור טבלת העונות:
נקח את ערכי העונות מטבלת ginf ונעשה distinct.

עבור טבלת המשחקים (Matches):
נקח את התאריך, שערי בית ושערי חוץ עבור כל משחק מטבלת ginf, נשייך FK של העונה מטבלת העונות ונשייך FK של קבוצת הבית וקבוצת החוץ מטבלת הקבוצות.

עבור טבלת האירועים:
נקח את זמן האירוע, סוג האירוע, מיקום הבעיטה, ואת סוג האירוע מטבלת events, נשייך FK של המשחק מטבלת המשחקים.

עבור טבלת נסיונות הבעיטה:
נקח את מיקום הבעיטה, האם זה גול, אמצעי הבישול, והסיטואציה מטבלת events ונשייך FK של האירוע מטבלת האירועים.