The background features a diagonal split between a light blue upper-left triangle and a black lower-right triangle. The entire page is decorated with a pattern of hexagons and interconnected lines. Some hexagons are solid blue, while others are black outlines. A network of thin blue lines connects small blue dots, creating a molecular or circuit-like structure. The title is positioned in the upper right, and the author information is in the lower right.

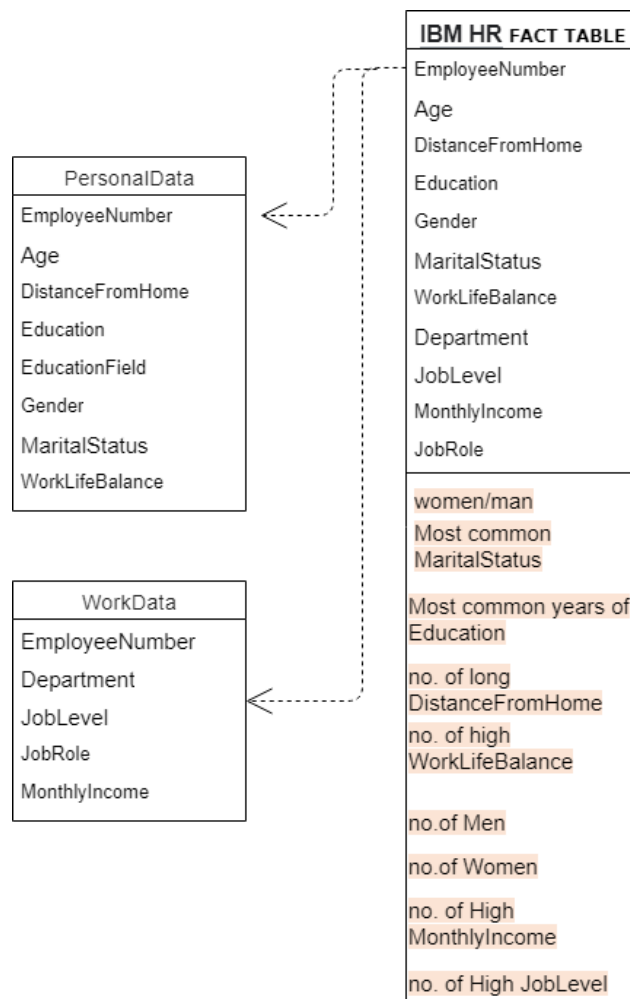
בינה עסקית – פרויקט חלק 2

מגישות
ורניקה פרידמן
לי קישון

סמסטר ב, 2022

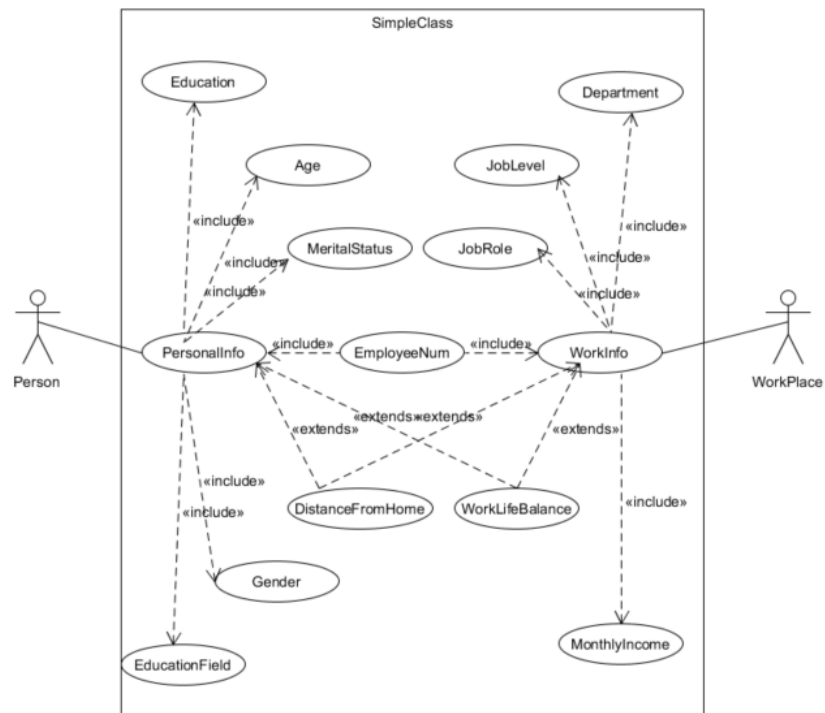
חלק א': הגדרת Data Warehouse

1. הסכמה שנבחרה היא סכמת STAR מאחר והיא מציגה את הנתונים שלנו בצורה הכי רלוונטית ומתאימה לתצורת המידע שלנו - דאטה מצומצם המתבסס על מספרי עובדים כPK. סכימות כוכב יצטרפו רק לטבלת העובדות עם טבלאות הממדים, מה שיוביל לשאילתות SQL פשוטות ומהירות יותר.
2. תיאור ה - Data Warehouse לאחר מימוש הסכימה ע"י תרשים ERD:



3. מימוש ה - Data Warehouse (מצורף בקובץ ה - EXCEL בגיליון DW)

4. use case :



5. Data Warehouse : מצורף בתיקייה.

חלק ב': הגדרה ומימוש ETL:

1. הגדרת תהליך ETL עבור אוסף הנתונים:

- Extraction - משיכת המידע מ - 2 הטבלאות כאשר הגורם המקשר הינו מספר העובד כאשר טבלה אחת בעלת הפרטים האישיים של העובד, והטבלה השנייה בעלת מאפיינים הקשורים לעבודה.
- Transformation - שיוך בין נתוני המידע של האדם ע"פ מספר עובד, כמו כן רידוד עמודות שאינן רלוונטיות לשאלת המחקר (EducationField) ורידוד נתונים חסרים.
- Loading - יצירת טבלה אחת מרכזת בעלת כל הנתונים הרלוונטים.

2. הגדרת תהליך ה- ETL Pipeline עבור אוסף הנתונים:

- Reference Data - הנתונים הרלוונטיים לשאלת המחקר שלנו הם:
 - גיל, מרחק מהבית, השכלה, מגדר ומצב משפחתי איכות חיים
 - Department, JobLevel, JobRole, MonthlyIncome
- Extract from Data Reference - המרת כלל הנתונים לקבצי CSV למטרת אפשרות למימוש בפייתון.
- Data Validation - וידוא הימצאות כלל הנתונים הנדרשים לטובת בדיקת שאלת המחקר.
- Transform Data - הוספת מדדים התורמים למדידת שאלת המחקר והפיכתם לבינארים.
- Stage - הזנת הנתונים ל - DW ניסיוני.
- Publish to Data Warehouse - ייצור הגרסה הסופית של ה DW.

3. מימוש תהליך ה- ETL Pipeline עבור אוסף הנתונים: מצורף בתיקייה.