# Junior Data Engineer zadatak

## zadatak

U file-u TRANSACTIONS\_HISTORY.csv nalaze se povijesni sales podaci za jednu trgovinu koja trguje tehnikom i uredskom opremom. U njemu se nalaze tri tipa podataka (datatype) koji se prepoznaju po prvoj koloni u retku:

* Transaction Header THDR, sa kolonama
  + Datatype (označava početak retka za THDR), string
  + TransactionID, string
  + OrderDate u formatu DD/MM/YYYY
  + TotalPrice, brojčana kolona
* Transaction Item TITM, sa kolonama
  + Datatype (označava početak retka za TITM), string
  + TransactionID, string
  + ProductID, string
  + ProductName, string
  + SoldPrice, brojčana kolona
* Product Catalog PCTL, sa kolonama
  + Datatype (označava početak retka za PCTL), string
  + ProductID, string
  + Category, string
  + Subcategory, string
  + ProductName, string

Potrebno je iz ove datoteke dobiti tri “čista” csv-a koji bi bili čitljivi bazi podataka, a svaki će sadržavati samo jedan tip podatka. Konačne datoteke neka se zovu THDR.csv, TITM.csv i PCTL.csv. Moraju sadržavati header s imenima kolona, a delimiter neka ostane pipe (znak |) kao i u originalnoj datoteci. Pretpostavi da će se ovaj algoritam vrtiti svaki dan na produkcijskoj okolini i da file može sadržavati značajno više redaka nego ovaj u primjeru.

Zadatak je poželjno napraviti koristeći Python i Pandas library. Rješenje se može dostaviti u obliku obične .py Python skripte ili .ipynb Jupyter notebooka i treba sadržavati sve korake od otvaranja početne datoteke do snimanja konačnih. Poželjno je kod popratiti komentarima koji će približiti tok misli.

## zadatak

Za sljedeći zadatak potrebno je koristiti sintaksu SQL Servera (TSQL) ili MySQL

Na slici je primjer dviju tablica u bazi.

EmployeeDetails

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **EmpId** | **FullName** | **ManagerId** | **DateOfJoining** | **City** |
| 121 | John Snow | 321 | 01/31/2014 | Toronto |
| 321 | Walter White | 986 | 01/30/2015 | California |
| 876 | Dean Smith | null | 18/06/2013 | California |
| 421 | Kuldeep Rana | 876 | 27/11/2016 | New Delhi |
| 986 | Anna Wang | null | 15/01/2014 | Toronto |

EmployeeSalary

|  |  |  |  |
| --- | --- | --- | --- |
| **EmpId** | **Project** | **Salary** | **Variable** |
| 121 | P1 | 8000 | 500 |
| 321 | P2 | 10000 | 1000 |
| 421 | P1 | 12000 | 0 |

Razmotri strukture tablica gore. Napiši upit koji će vratiti

1. Sve zaposlenike i njihove plaće, čak i ako ne postoji podatak o plaći za zaposlenika
2. Sve zaposlenike koji su ujedno i menadžeri
3. Menadžere i projekte na kojima su involvirani
4. Drugog najplaćenijeg zaposlenika po svakom gradu