Project Future

6ος ΚΥΚΛΟΣ

Project Assignment

Business Intelligence & Data Sciences

Project Future 6

Powered by

Code.Hub

ΤΡΑΠΕΖΑ ΠΕΙΡΑΙΩΣ

Η ευκαιρία σου να ξεχωρίσεις, εδώ!
REGENERATION®
BY GLOBAL SHAPERS ATHENS HUB

Contents

StayMore
Short-term renting app

Introduction

This integrative group project aims at encouraging students to apply the knowledge and experience learned in the class towards a real-life business intelligence system. StayMore is a (fictitious) property management company which operates in a certain city. The administration office already has a custom CRM system with a database server and the functionality to provide csv files. In terms of data content, the StayMore office provides information about house listings, rental calendars, customer reviews. Your own project will focus exclusively on the extraction, transformation, management to create the OLAP system. Furthermore, data visualization, intelligence gathering, and presentation of the company data will be organized and implemented.

Project Scope and Deliverables

For your team project, you will need to collect all the existing company data in its current state and reorganize it in an efficient way to handle and use for reporting and analytic purposes. The scope of the project includes the following broad objectives:

- Create a centralized view of the data, so that by having immediate access to all company data will improve their customer interaction and decisionmaking process.
- Create an integrated business intelligence system that satisfies the user requirements as stated further below in this document.

- Create data mining and visualization arrangements to help to answer faster and more accurately questions.

Your project deliverables which will support the aim are identified as deliverables D01-D11 in the following sections. You will collect all deliverables and submit them as your project portfolio work.

Overview of Project Work

For running this project, you are advised to frequently meet as a team, and discuss and agree on your implementation plan and actions. This means that you must end up with a clear understanding of

a. the roles and responsibilities of the team members.

b. the project requirements.

c. the data requirement.

d. the way you will run your project.

e. the tools you will use for the technical work.

f. the tools you will need for the running of your team.

g. the deliverables of your work You will use some of the above decision content in the deliverables outlined next.

Data Description

The company's dataset represents house listings, rental calendars and customer reviews.

- House listings: detailed house listings data including full house descriptions and average review scores.
- Rental calendar: detailed rental calendar data for house listings, including listing id, rental price and availability for each calendar day.
- Customer reviews: detailed review data for house listings including a unique id for each reviewer and their detailed comments for the property.

Note: The actual datasets are provided by Airbnb on a Creative Commons CC0 1.0 Universal (CC0 1.0) "Public Domain Dedication" license, so it is free to use in this project. For the sake of our scenario, we will assume that it is the data of the StayMore offices.

The data are given in csv files and as a back-up file from an OLTP database. The files are available in a folder called project-data-files in the teams' application.

Problem Definition and Business Intelligence Scope

For initiating your project work, you are asked to consider the problem the company faces, the current business landscape and the proposed solution. Based on this analysis, prepare the following deliverables:

D01. Problem and proposed solution overview.

D02. Analysis of the problems and needs that will be covered with the proposed BI solution.

D03. Team organization, tasks identified and project timeline.

You may use any kind of tools, reporting style and diagramming techniques to create the above deliverables. The result must be such that a reader with no prior knowledge of the scenario and the project will understand what you will deliver and why.

System Design & Development

To progress with your project, you are advised to include the following steps in your process.

- Create a staging environment for the necessary data.
- Define, design and implement the necessary Extract, Transform, Load (ETL) tasks.
- Design and implement a data warehouse and necessary data marts (if needed).
- Define and implement necessary Information Visualization & Data Analytics and Dashboards.
- Discuss the implemented Business Reports and corresponding findings Any DBMS and BI infrastructure/application may be used for the above steps. SQL Server, Power BI and Python are the recommended tools, platforms and technologies for this project.

Once the tasks are complete, you need to supply the following deliverables:

D04. List of development and collaboration tools used.

D05. BI system overview, architecture & system workflow/workflows.

D06. Database diagrams and source code (OLTP, Staging, DW, DMs).

D07. Data flow diagram/diagrams & transformations explanations

D08. Code files used for data processing.

D09. Data Visualization platform's (e.g., from Power BI) files, relevant screenshots and description of functionality and intended usage and purpose (see also paragraph 7).

Data Visualization & Advanced Analytics

The core of the project is to perform a series of analytic calculations and visualizations. These results will be in a live dashboard for real time presentation.

D10. A series of analytics can provide answers to questions like:

a. Which are the company's top priced properties

b. Which are the company's top-rated properties

c. Which hosts are the busiest?

A secondary objective is to build a machine learning model that predicts the price of a listing depending on its attributes (e.g., number of bedrooms, property type, etc.). This model will be used for suggesting the price that a new listing should take. Then you should use this house-price model to answer the following questions:

d. Which attributes are the most critical to the price of a listing?7

e. Which listings are the most undervalued?

f. Which listings are the most overvalued?

g. Identify candidates based on their value and ratings whose hosts could be notified for increasing their price.

Notes:

● *The dataset needs to be properly pre-processed (cleaning, encoding, etc.) in order for the model to be trained.*

● *Be careful of which features you feed the model. Some of them may not be available in a new listing. Your deliverables will include D10. A description and presentation of the analytics along with the source code and other necessary files.*

Project Evaluation

Along with your work on the project itself, you are asked to prepare an evaluation of your resulting system. This will be part of your submission to the StayMore company, so that they are fully aware of the system's strengths, weaknesses, and mode of operation.

Include the following documentation in your work deliverables:

D11. Project Evaluation, advantages, limitations, and suggestions for future steps.

Project deliverables

You will need to collect the deliverables D01-D11 named in this document and submit them as your team's final deliverable – this will be your team's project portfolio. You can use the files tab of the private channel of your team in Microsoft Teams. You will also need to prepare a presentation on your project work. This is a presentation that you will give as a team at the end of the course. Details on the content will also be discussed during the contact sessions.

Appendix

	A	B	C	D	E	F	G
1	listing_id	date	available	price	adjusted_price	minimum_nights	maximum_nights
2	2818	10/10/2020	f	$59.00	$59.00		
3	2818	10/11/2020	t	$59.00	$59.00	3	1125
4	2818	10/12/2020	t	$59.00	$59.00	3	1125

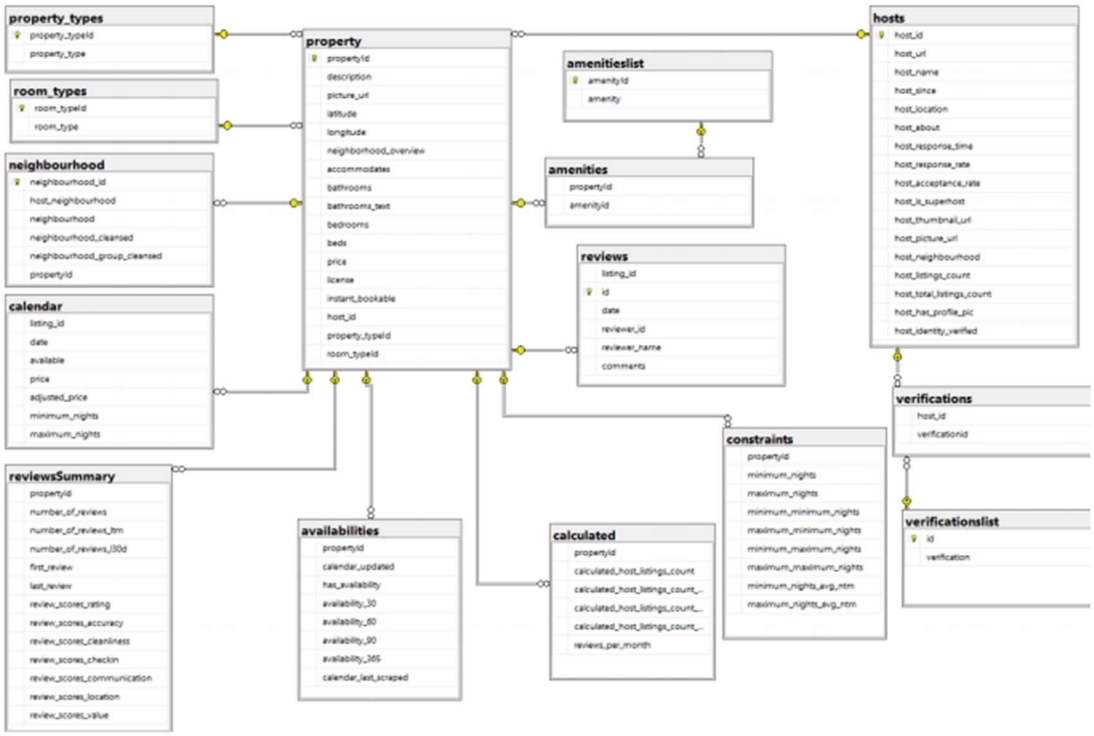
Image 1.a. Calendar.CSV Image

	A	B	C	D	E	
1	listing_id	id	date	reviewer_id	reviewer_name	comments
2	50904	31511792	5/6/2015	19482395	Jihae	Karinâ€™s â€œAplaceâ€ is ab
3	116134	972063	3/5/2012	928644	AurÃ©lien	Amazing flat, really close from

1.b. Reviews.CSV Image

	A	B	C	D	E	F
1	id	listing_url	scrape_id	last_scraped	name	description
2	50904	https://www.airbnb.com/rooms/50904	20201026165813	10/26/2020	aplace/antwerp: cosy suite - fashion district	Decorated in a vintage style combined with a fe
3	116134	https://www.airbnb.com/rooms/116134	20201026165813	10/26/2020	Trendy Vacation Apartment Antwerp	Enjoy your stay at our 4 person apartment in the

1.c. Listings.CSV Image



2. The database diagram