Property Analysis BI Developer- Competition Tasks - Part 1

Introduction:
Welcome! The Property Analysis Competition Task is designed to check the potential of each and every intern before they actually move into the project. Also to check if you can work with a team spirit and achieve the goals before the deadlines. While the instructions tell you what to do, they do not always necessarily tell you how to do it. This is deliberate as it is important for you to develop an independent drive to solve problems on your own. You should have a good idea where to start from your 6 weeks of training and the on-boarding tasks.

MVP Studio has an opening for an eager BI developer who is keen to be involved in setting up our business intelligence capability from the beginning, for providing insights for property investors and other business users.

Competition Scenario:

Part 1

1. Transform data from multiple data sources, ensure the accuracy and integrity of relevant data sets.
2. Design, build and test of a data warehouse solution.

Part 2

3. Access, analyse and interpret data to deliver easy to consume and actionable insights for the business with a dashboard report.

Note: Download the raw datasets from this link. Once you have completed your tasks, upload your solution files to a folder in your Google Drive then share its link to assigned Mentor.

Task 1: Standardising raw data sets and data modelling.

a) Data Standardisation is a critical process that transforms data from different sources to a common format so that they are consistent with each other and ready to be integrated to their destination systems for further analysis. This task involves steps such as understanding and cleaning your data, dealing with missing values, matching data types, combining or creating new columns / rows in a table, etc. It can be done with any tools that come in handy, such as Excel and PowerBI.

b) Some examples for standardised datasets are listed below. You can create your own standardised formats. Upload your standardised datasets in your submission folder:

	A	B	C	D	E	F	G	H
1	StateName	StateCode	City/Town	Postcode	Suburb	PropertyMedianValue	UpdatedYear	UpdatedMonth
2	New South Wales	NSW	Sydney	2000	Dawes Point	1044350	2017	6
3	New South Wales	NSW	Sydney	2000	Haymarket	1044350	2017	6
4	New South Wales	NSW	Sydney	2000	Millers Point	1044350	2017	6
5	New South Wales	NSW	Sydney	2000	Parliament House	1044350	2017	6
6	New South Wales	NSW	Sydney	2000	Sydney	1044350	2017	6
7	New South Wales	NSW	Sydney	2000	Sydney South	1044350	2017	6
8	New South Wales	NSW	Sydney	2000	The Rocks	1044350	2017	6
9	New South Wales	NSW	Sydney	2007	Broadway	1051540	2017	6

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	school_id	postcode	school_name	school_address	town_suburb	school_lat	school_lon	student_number	ICSEA_Value	level_of_schooling	school_specialty_type	school_subtype	school_gender	state_code
2	1001	2046	Abbotsford Pub	350 Great North R	Abbotsford	-33.853	151.13	602	1127	Primary Schoo	Comprehens	Kinder to Yea	Coed	NSW
3	1002	2336	Aberdeen Publi	Segenhoe St	Aberdeen	-32.166	150.89	241	917	Primary Schoo	Comprehens	Kinder to Yea	Coed	NSW
4	1003	2326	Abermain Publi	Goulburn St	Abermain	-32.809	151.43	264	887	Primary Schoo	Comprehens	Kinder to Yea	Coed	NSW
5	1007	2629	Adaminaby Pub	9 Cosgrove Street	Adaminaby	-35.993	148.78	17	973	Primary Schoo	Comprehens	Kinder to Yea	Coed	NSW
6	1008	2289	Adamstown Pul	Bryant St	Adamstown	-32.932	151.73	263	1024	Primary Schoo	Comprehens	Kinder to Yea	Coed	NSW
7	1009	2729	Adelong Public	Gilmore St	Adelong	-35.312	148.06	81	966	Primary Schoo	Comprehens	Kinder to Yea	Coed	NSW
8	1015	2527	Albion Park Pub	Tongarra & Hamil	Albion Park	-34.57	150.77	435	977	Primary Schoo	Comprehens	Kinder to Yea	Coed	NSW
9	1016	2340	Timbumburi Pul	542 Kia Ora Lane	Timbumburi	-31.204	150.92	144	995	Primary Schoo	Comprehens	Kinder to Yea	Coed	NSW

	A	B	C	D	E	F	G	H	I	J
1	stop_id	Station_Suburb	Street_Name	Street_Type	Entrance_Type	LAT	LONG	Exit_Numl	Mode	StateCode
2	1	Aberdeen	Macqueen	St	Stairs	-32.1669	150.892		Train	NSW
3	2	Aberdeen	Macqueen	St	Ramp	-32.1669	150.892		Train	NSW
4	3	Adamstown	Park	Ave	Path	-32.9337	151.7205		Train	NSW
5	4	Adamstown	St James	Rd	Stairs	-32.9334	151.7204		Train	NSW
6	5	Adamstown	St James	Rd	Ramp	-32.9334	151.7203		Train	NSW
7	6	Adamstown	Park	Ave	Path	-32.9338	151.7202		Train	NSW
8	7	Albion Park	Burroo	St	Ramp	-34.5634	150.7994		Train	NSW
9	8	Albion Park	Princes	Hwy	Ramp	-34.5635	150.7992		Train	NSW
10	9	Albion Park	Princes	Hwy	Stairs	-34.5629	150.7986		Train	NSW
11	10	Albion Park	Princes	Hwy	Ramp	-34.5628	150.7984		Train	NSW

c) Design a simple data warehouse for the data given above using Kimball dimensional modelling techniques. You may need to create additional tables (e.g. DimDate, DimSuburb) to meet the requirements. Draw a diagram for your data warehouse schema using http://draw.io or a similar tool (submit as a screenshot).

Task 2: Based on your data model, create a database (data warehouse) with its tables to store your data in SSMS using your Windows authentication.

- Generate and submit a copy of SQL script (*.txt or *.sql) after database and tables are created in your SQL Server.