CODE BASICS

INTRODUCING

AI ENABLED DATA ANALYST

DA 2.0

AI Proof
Data Analyst

Chapter 2: Using AI tools for technical work

1. Data Collection On MySQL and Excel Using ChatGPT:

Prompt 1:

I have MYSQL database with following tables,

1) movies: columns: movie_id, title, language_id

2) languages: columns: language_id, name

Can you write SQL query to join these two so that I get all the columns in movies table and also the language name?

Prompt 2:

I've two excel files.

1) movies.xlsx: columns: movie_id, title, language

 2) financials.xlsx: columns: movie_id, revenue, budget movie_id is a unique common column.

Can you tell me how can I merge them using VLOOKUP such that I get a single file with all the columns?

2. Data Collection Using ChatGPT in Python:

Prompt 1:

I have a CSV file named fact bookings.csv located in the datasets folder. How can I read it and store it in a variable called of bookings ? Also, I would like to perform a sanity check by displaying the top 5 rows.

Prompt2:

Could you assist me with these preliminary questions using Pandas code?

1. How can I determine the number of rows and columns in the df_bookings DataFrame?

2. What are the unique categories present in the room_category and booking_platform columns of the df bookings DataFrame?

3. How can I find the number of rows corresponding to each category in the booking_platform column of the df_bookings DataFrame? Additionally, could you assist me in creating a bar chart to visualize this information effectively?

3. Power BI: Data Cleaning:

Prompt 1:

In Power Query, I have a column named "property_name" that contains values like "Atliq Bay" and "Atliq bay" in different formats. My goal is to standardize the format of these values to "Atliq Bay". How can I achieve this?

Prompt 2:

In Power Query, I have a column named "property_type" that has leading space characters around the categories. I need to remove these unwanted characters. How can I accomplish this in Power Query?

Prompt 3:

In Power Query, I have a column named 'city|city_code' where the city and city_code values are separated by the delimiter '|'. I want to split this column into two separate columns. How can I accomplish this in Power Query?

4. SQL: Data Cleaning

Prompt:

In my MySQL database, I've a table called product sales. Write a query to retrieve last year's data. Filter any data where quantity is negative or the order_price is O or less. Column names are: product_id, product_name, quantity, order_price, order_date

5. Python: Data Cleaning

Prompt 1:

How can I remove the invalid guests (those containing 0 or less than 0) from the no guests column in the df bookings DataFrame and update the same DataFrame with the cleaned data?

Prompt 2:

How can I fiter the data in the df bookings DataFrame based on the revenue_generated column, keeping only the data that falls within 3 standard deviations from the mean, and update the same DataFrame with the cleaned data?

Prompt 3:

How can I filter the data in the df_age_bookings dataframe where the value of successful bookings is less than equal to the capacity and update the same DataFrame with the cleaned data?

6. Python: Data Transformation

Prompt:

How to create a new column named occ_pct in the df_agg_bookings dataframe? This new column should contain the ratio of successful_bookings to capacity, represented as a percentage with two decimal values.

7. SQL: Data Transformation

Prompt 1:

We have a MYSQL table called orders. I want to analyze all the transactions in US $ transactions. The problem is in the table they are stored in non uniform format such as "USD, "US$", "Dollars". You can also have other currencies. How do I get all the records where the currency is mentioned in these 3 different formats?

Prompt 2:

Can you change it to match exact texts?

8. Power BI: Data Transformation

Prompt 1:

In PowerBI, How do I create a new calculated column in the bookings_data table called "Availability Status" that returns "sold out" if the value in the "successful_bookings" column is equal to the value in the "capacity" column, and "vacant" otherwise?

Prompt 2:

In PowerBI, How do I create a new calculated column in the bookings_data table called "occ%" that calculates the percentage of successful_bookings to the capacity. The percentage should be rounded to two decimal places.

9. Data visualize using ChatGPT:

Prompt:

How to do the below scenario in Pandas?

I have a dataframe named df bookings with columns such as booking id, property id, booking date, check_in_date. checkout date, no guests, room category, booking platform, ratings given, booking status, revenue generated, and revenue_realized. My objective is to group the data by booking platform and analyze the revenue generated by each platform, Additionally, I would like to create an appropriate chart to visualize the output data.

10.   General Use Cases:

1. Explain existing code:

Prompt 1:

Can you explain me this SQL Code:

```
CREATE PROCEDURE get_top_n_products_by_net_sales(

        in_fiscal_year int,

        in_top_n int

        )

        BEGIN

    select

        product,

        round(sum(net_sales)/1000000,2) as net_sales_mln

    from gdb041.net_sales

    where fiscal_year=in_fiscal_year

    group by product

    order by net_sales_mln desc

    limit in_top_n;

        END
```

Prompt 2:

Can you explain me this Pandas code:

```
def get_threshold(row):

   occ_per = row['occ%']
```

```python
    if occ_per > 80:

        return "Good"

    elif occ_per > 50 and occ_per <= 80:

        return "Average"

    else:

        return "Bad"

df_agg_bookings['threshold'] = df_agg_bookings.apply(get_threshold, axis = 1)
```

Prompt 3:

Can you explain me this PowerBI code:

Revenue WoW change % =

Var selv = IF(HASONEFILTER(dim_date[wn]),SELECTEDVALUE(dim_date[wn]),MAX(dim_date[wn]))

var revcw = CALCULATE([Revenue],dim_date[wn]= selv)

var revpw = CALCULATE([Revenue],FILTER(ALL(dim_date),dim_date[wn]= selv-1))

return DIVIDE(revcw,revpw,0)-1

2. Write Code:

Prompt:

Select all the Hollywood movies released after the year 2000 that made more than 500 million dollar profit. The unit in a database record can either be a million or a billion. If the unit is in billion, convert it to million.

Tables with respective columns:

movies: movie_id, title, industry, release_year, imdb_rating, studio, language_id

financials: movie_id, budget, revenue, unit, currency

3. Debugging Code:

Prompt:

Can you find out the error in this code and provide the correct query.

```sql
WITH product_if_target AS (

    SELECT

        p.product_name,

        SUM(CASE WHEN f.in_full = 1 THEN 1 ELSE 0) AS if_count,
```

```
      COUNT(f.order_id) AS total_count

   FROM

      fact_order_lines f

      JOIN dim_products p ON p.product_id = f.product_id

   GROUP BY p.product_name

)

SELECT

   product_name,

   ROUND(if_count / total_count * 100), 2) AS IF_percentage

FROM

   product_if_target

order by IF_percentage DESC;
```

4. Generating Sample Dataset:

Prompt:

can you create a sample dataset with these columns:

columns:

order_id,   order_placement_date , customer_id, product_id,   order_qty,
        agreed_delivery_date, actual_delivery_date,   delivery_qty

More additional information about the columns:

1. order_id: Unique ID for each order the customer placed

2. order_placement_date: It is the date when the customer placed the order

3. customer_id: Unique ID that is given to each of the customers

4. product_id: Unique ID that is given to each of the products

5. order_qty: It is the number of products requested by the customer to be delivered

6. agreed_delivery_date: It is the date agreed between the customer and Atliq Mart to deliver the products

7. actual_delivery_date: It is the actual date Atliq Mart delivered the product to the customer

8. delivered_qty: It is the number of products that are actually delivered to the customer

Make sure to give output in tabular format and give 10 rows

Chapter 3: Using AI tools for Communication

1. Saying things in a compact way:

Prompt:

Say this sentence in a more compact way using only 5 words

Can you fill my gas tank with full with regular petrol. And yes I will use my credit card for the payment.

2. Asking clarifying questions on requirements:

Prompt:

Make below email concise. Make bullet points for questions that are being asked to Wanda

Hello Wanda,

It was a nice call this morning for project kick off. I have some follow up questions post our project kick off meeting, I hope you can answer those. In terms of the dashboard we are building are we going to include building the market share visual in phase 1? Who can give me the required data for the past performance of our PPC campaigns, ROAS, marketing spend etc... you know I am bit confused and not sure what to do so asking for your help... Also I am attaching the roug project proposal here in this email so please kindly look at it and thanks very much once again for such a detailed meeting today morning.

Thanks you.

Regards,

Peter Pandey

3. Responding to Questions

Prompt:

Make below email concise. Make bullet points for Answers

Hello Peter,

Thank you for attending the call today morning. I also enjoy talking about requirements in the details and I hope you are doing well by the way. Regarding your questions, see for the market share visual, let's see! I am thinking we should include it in phase 1 because this will bring transparency to our marketing team. The answer of your second question is that you can contact Lauki Lal who is our data engineer. He knows about data catalog, handles our SQL database. By teh way he is a talented guy.He will help you get required data such as PPC campaign, ROAD and such.

Thanks very much,

-Wanda

4. Making tough communications easier

1. Dealing with a manager:

Prompt:

My manager is giving me very tight, unreasonable deadlines. My mother's health is not good and I need to be home before 6.30 pm but when I leave at 6 pm sharp he thinks I am not hardworking. How do I communicate my situation through emall to my manager such that he understands! Emall should be less than 10 lines

2. Appraisal/performance meetings:

Prompt:

I have an appraisal meeting in 2 weeks, my performance is great and I deserve 20% hike.

Let's say my manager says she can give me only 8% hike, give me some real conversation examples of how I can negotiate

These are some areas I did well

-Completed project athens successfully

- onboarded a new team member and made her successful

-worked on feedback regarding communication skills from my previous review meeting and improved it

3. Not able to meet a commitment:

Prompt:

I have a task deadline in 4 days and looks like I won't be able to meet it.

The real reason is I was careless and did not manage my tasks well. But I'm willing to make some extra effort and complete it with a quality check within 2 additional days. My manager asked me last week If I will be able to deliver on time and I said Yes.

How can I communicate about this scenario professionally over an email in about 150 words to my manager in such a way that she will still see me as a credible person?

5. Ask for enhancement tips

Prompt:

I'm presenting some insights to my business stakeholders on sales growth over last 12 months and giving them few suggestions from my end.

How can I start the presentation so that I grab their attention instantly?

Give me 3 tips

6. Preparing Presentations

Prompt:

Below is the story of how EY used microsoft power platform to speed time to market. Please create 5 slides power point presentation for this

Business travel and living abroad for work can be exciting career milestones—and also require a lot of legwork to coordinate changes in tax status, compensation, and immigration. To centralize that coordination for its clients, EY teams created the EY Mobility Pathway (EYMP). Despite the complexities and massive amounts of data involved, the firm launched EYMP quickly using Microsoft Power Platform, Dynamics 365, Azure, and a custom React portal. EY teams brought the product to market in record time, delivering a better user experience and value-added services to its clients.

We used Microsoft Power Platform, and within six weeks, we built a prototype that we could demo to EY clients.

Michael Maloney: Senior Technical Architect

Ernst & Young LLP.

Centralized mobility platform

Companies around the world know the EY organization for helping clients build a better working world, foster trust in the capital markets, and create long-term value for society. Working across assurance, consulting, law, strategy, tax, and transactions, EY teams ask better questions to find new answers for the complex issues facing our world today.

Many of its clients turn to EY teams for enterprise mobility support. EY professionals helps them manage everything related to corporate international travel or relocation, such as employee immigration, tax implications, compensation, visas, and lodging.

The EY organization has been offering enterprise mobility services for few years, but clients and staff had to handle each aspect of these services individually. The firm began to develop a product in-house called the EY Mobility Pathway (EYMP) to consolidate the web of legacy systems. It was incredibly complex and involved huge amounts of data.

Rapid development unlocked

EY teams have worked closely with Microsoft for many years and won the 2020 US Microsoft Partner of the Year Award for Advisory Services. In early 2019, the EYMP team turned to Microsoft Power Platform to create a proof of concept for the EYMP. "We used Microsoft Power Platform and within six weeks, we built a prototype that we could demo to EY clients," says Michael Maloney, Senior Technical Architect at Ernst & Young LLP.

EY developer teams determined that all case management could take place in the Microsoft Power Platform. A "case" refers to the employee that the mobility services focus on, and each aspect of a case is called a task. Internally, EY analysts and case coordinators use EYMP to help move client cases forward. Clients also access EYMP, including the employee going overseas, who might need to upload relevant documents, and the HR staff or corporate managers who track the employee's case. Developers augmented EYMP with Microsoft Power BI for reporting and analytics, which the HR and corporate managers access to gain visibility into their overall case status across various tasks.

EYMP uses Microsoft Dataverse to store data related to the cases, such as citizenship, contact, and dependent information. The developers also took advantage of Microsoft Power Automate to automate the many tasks within a case. EYMP automatically creates cases when triggered by certain conditions and generates tasks as the case moves through different stages. "We've added a lot of automation with Power Automate as our workflow engine," says Maloney. "We've drastically reduced manual work, even on complex workflows, which is fantastic."

EY developers used React, which its creators describe as "a JavaScript library for building user interfaces," to create a custom front-end portal. React communicates directly with the various back-end systems through APIs. "We trained all our developers in using Microsoft Power Platform because it's extremely powerful and flexible," says Maloney. "It's been easy for our developers to work across Microsoft Power Platform with the React portal just by adding APIs."

To manage traffic to the EYMP portal, EY teams uses Microsoft Azure Application Gateway. The organization also relies on Microsoft Dynamics 365 Customer Service to support case management, including document management and contact access for customer profiles.

Unprecedented speed to market

EYMP is an incredibly complex solution, but EY teams were able to take it to market quickly by taking advantage of Microsoft Power Platform. The team began development in May 2019 and took EYMP live with the first client in November of the same year. "We built our product quickly with Microsoft Power Platform because it's a low-code, highly configurable platform that handles large datasets," says Doug Fighter, Senior Manager at Ernst and Young LLP . "The EYMP leadership team was happy with the speed with which we built this product."

Added value to clients

Through EYMP, the organization now offers a more streamlined, sophisticated experience to its clients to help them manage corporate travel. "The biggest impact for EY clients is the seamless experience offered by EYMP," says Fighter. "Rather than information in siloed applications, we now provide visibility and a rich user experience for EY clients to help manage information within one tool. It's a huge upgrade that clients are thrilled with."

With all enterprise mobility services in one cloud-based platform, EY teams have positioned itself to serve the unique needs of large clients more easily. "EYMP is a supporter for some of the enterprise clients. We now can deploy into an environment specific to them and customize EY services for their needs in just 2 months instead of 6 to 12 months," says Fighter.

He concludes, "We have the flexibility and the business logic to manage thousands of different combinations of country-to-country travel within one tool, which has made a real difference for EY clients and our internal teams. Choosing Microsoft Power Platform has definitely been a win-win for us all."

Chapter 4: AI for Interview help

1. Using AI to search for job titles

Prompt:

I am looking for a data analyst job. I know SQL, Excel, Power BI. What kind of job titles should I search for?

2. Using AI to customize resume

(Video reference: https://codebasics.io/courses/bootcamp/1/Building-Online-Credibility-Portfolio-Projects-ATS-Resume/lecture/1257)

Prompt 1:

Tailor my resume as per the job description. Job description and resume are provided below,

Job Description

Resume

Prompt 2:

Fit in on page with more emphasis on Power BI

3. Asking for referral

Prompt 1:

I'm applying for a job at a company called Fractal and found someone who is working there on Linkedin. Can you help me write a linkedin message asking for referral in not more than 100 words.

I will give 3 Inputs based on which you give the referral message

Input 1: Person's about me section from whom I want referral

Input 2: Job Description

Input 3: My Resume

Prompt 2:

Can you shorten this text to 5 lines and adjust the template to add project and resume links

4. Provide Feedback For Better Response

Prompt:

Can you make the subject line look more unique

5. Mock Interviews

1. Questions related to a job posting:

Prompt 1:

I am going to give an interview for a power bi developer position. Ask me some interview questions based on the below job description.

Job description:

Minimum 3 years of Strong working experience in MS Power BI tool, Data Analytics and Data Visualisation

Atleast 1 year experience in Python Programming

Power BI Cloud / Workspace

Power Apps

Power Automate

Power Insights

Should have working experience on Creating Gateways and API's connection to Power BI

Knowledge on Power BI Embed (Desirable)

Should possess strong knowledge on DAX / MDX Query languages

Expert level knowledge in ETL tools, Creation of Database using Excel / SQL / any other applications if required (Desirable)

Experience working with SSAS – Tabular Model or Multi-dimensional models

Knowledge on SSMS and SSIS systems or any other analytical tools – Desirable

Knowledge on database development / management (SQL / ORACLE / Cloud based database) would be added advantage (Desirable)

Working knowledge of MS Office, relational databases and ERP systems (e.g. SAP)

Worked on complete SDLC in Power BI development projects

Preferable to have experience in Python and R development

Analytical mind with business acumen, Problem-solving aptitude

Good communication skills (Mandatory)

Must have customer handling experience

Prompt 2:

Ask these questions one by one

2.Questions related to your resume:

Prompt:

I am interviewing for a data analyst position. Ask me resume and projects related questions based on below resume.

Resume

----

6. Technical Interview

1. SQL interview: Basic one

Prompt 1:

Conduct my SQL interview for a data analyst position

Ask questions one by one

Prompt 2:

Now ask me to write some SQL queries

2. SQL interview: Only Queries & Rating

Prompt:

Take my SQL interview for a data analyst position. Ask only one question at a time. Ask me to write queries only, no theoretical questions please.

3. SQL interview: Difficulty Level

Prompt:

Take my SQL interview for a data analyst position. Difficulty level should be HIGH. Ask only one question at a time.

4. Mixed Interview: Excel, SQL, Power BI

Prompt:

Conduct my technical interview for a data analyst position. Ask questions on Excel, Power BI and SQL

5. Power BI Interview: Only Formulas

Prompt:

take my power bi interview. ask questions on writing DAX formulas for a given use case

7. Behavioral / Case study

Prompt 1 :

Conduct my Interview for a data analyst position. Ask case studies type questions, one by one

Prompt 2 :

Please conduct a behavioral mock up interview with me, ask at least 5 questions.

I will give 3 Inputs based on which you ask questions

Input 1: Company Overview from theri Linkedin page

Input 2: Job Description

Input 3: My Resume

Please ask behavioral questions, one by one.

Prompt 3 :

Ask some behavioral questions that are asked in data analyst interview conducted by a company called <Company Name>

8. Connecting with people on LinkedIn

Prompt:

I'm a data analytics aspirant & want to create a positive impression with Dhaval Patel, founder of Codebasics. Write me a LinkedIn connection message in 100 words. This is his about me section

https://www.linkedin.com/in/dhavalsays/