KKR 2nd Interview

Excel

Data cleaning

- Text
  - Remove/replace text or Create categorical variable: select the region, Find & Select -> Replace ; replace (*): everything within (xxxxxx)
  - Combines texts from multiple ranges and/or strings: CONCAT(text1, " ", text2, ...) ; TEXTJOIN(delimiter, ignore_empty, text1, [text2], ...) with a range
  - Return characters in a text string: LEFT(text, [num_chars]); RIGHT(text,[num_chars]); MID(text, start_num, num_chars)
  - Convert the case for texts: UPPER(text); LOWER(text); PROPER(text)
  - Split texts: Data -> Text to Columns ; TEXTSPLIT(text, col_delimiter, [row_delimiter])
  - TEXTBEFORE(text, delimiter, [instance_num], [match_mode], [match_end], [if_not_found]) : returns text that occurs before a given character or string
  - TEXTAFTER(text, delimiter, [instance_num], [match_mode], [match_end], [if_not_found]): returns text that occurs after a given character or string
  - Search & replace: SEARCH(find_text, within_text, [start_num]) ~ return position; REPLACE(old_text, start_num, num_chars, new_text); SUBSTITUTE(text, old_text, new_text, [instance_num]); TRIM(text) ~ remove all the spaces
- Format
  - Identify duplicates: Home -> Conditional Formatting -> Highlight: Duplicate Values ; or use IF(XX=XX, "Duplicate", "")
  - Remove duplicates: Data -> Remove Duplicates
  - Fit width/height: select the region -> Home -> Format -> AutoFit Height/Width
  - Fill the blank with NA: select the region -> Find & Select -> special -> blank -> type "NA": ctrl enter
  - Fill the blank with the cell above: select the region -> Find & Select -> special -> blank -> point to the cell above: ctrl enter
  - Convert a row/column to an array: WRAPCOLS(vector, wrap_count, [pad_with]); WRAPROWS(vector, wrap_count, [pad_with])
  - Deal with error: IFERROR(function, "NA")
  - Data -> Flash Fill: automatically fills your data when it senses a pattern.
- Date/Time
  - Difference in date: DATEDIF(start_date, end_date, unit), unit = "y" or "m" or "d"

    - Date +/- : EDATE(start_date, months); date +/- (days)
    - Time +/- TIME(hour, minute, second)
    - Text to date: DATEVALUE(date_text)
- Consolidate data
    - VSTACK(array1,[array2],...) : appends arrays vertically and in sequence to return a larger array
    - HSTACK(array1,[array2],...) : appends arrays horizontally and in sequence to return a larger array
    - TOCOL(array, [ignore], [scan_by_column]) : returns the array in a single column
    - TOROW(array, [ignore], [scan_by_column]) : returns the array in a single row
- Power Query
    - Cleaning: remove duplicate/error, change data type, transform value(rounding, log, power...), split/merge columns, groupby, format(upper, trim...)
    - Combine multiple worksheets: open a new workbook -> get data: transform data -> keep only the source -> remove redundant columns -> expand sheets
- Macro
    - Developer -> Record Macro -> ... -> Stop Recording -> go to target cell(s) -> Macro: run
    - VBA: https://www.wallstreetmojo.com/vba-examples/ ; https://cheatography.com/tag/vba/

Data Analysis

- Statistics
    - Basic functions: MAX/MIN/LARGE/SMALL/AVERAGE/MEDIAN/MODE(numbers/names/arrays,[...]) ; RANK(number, ref, [order]) ; FREQUENCY(data_array, bins_array) ; PERCENTILE(array,k) ; STDEV(number1,[number2],...)
    - Data -> Data Analysis -> Descriptive statistics
    - Data -> Data Analysis -> Correlation Analysis
    - Data -> Data Analysis -> Linear Regression
- Sum & Count
    - SUMIFS(sum_range, criteria_range1, criteria1, [criteria_range2, criteria2], ...) : adds all of its arguments that meet multiple criteria
    - SUMPRODUCT(array1, [array2], [array3], ...): returns the sum of the products of corresponding ranges or arrays
    - Conditional SUMPRODUCT : eg. SUMPRODUCT(IF(C5:C16="red",1,0),D5:D16,E5:E16)
    - COUNT(single_column_table) counts the number of records that contain a number in a single-column table.
    - COUNTA(single_column_table) counts the number of records that aren't *blank* in a single-column table

    - COUNTIFS(criteria_range1, criteria1, [criteria_range2, criteria2]...) applies criteria to cells across multiple ranges and counts the number when met
- Insert -> Pivot Table
    - Insert -> sort/filter/group..., field(sum, average, count...), show value as % of total, ... => double click the cell to see the filtered original data table
    - Dynamic filter: PivotTable Analyze -> Insert Slicer -> select categories (can overlap); right-click Slicer -> Report Connections to link to 1+ tables
    - Dynamic timeline: PivotTable Analyze -> Insert Timeline (date)
    - Dynamic chart: select the pivot table -> insert charts: can dynamically adjust based on the Slicers
    - Layouts: (Design -> Layouts)subtotals, grand totals, blank rows... ; (PivotTable Analyze -> Show) ...
- Table functions
    - VLOOKUP(lookup_value, table_array, col_index_num, [range_lookup]) : find things in a table or a range based on column
    - HLOOKUP(lookup_value, table_array, row_index_num, [range_lookup]) : find things in a table or a range based on row
    - XLOOKUP(lookup_value, lookup_array, return_array, [if_not_found], [match_mode], [search_mode]) : find things in a table or range by column
    - INDEX(array, row_num, [column_num]) returns value based on location ; MATCH(lookup_value, lookup_array, [match_type]) returns the location
    - => INDEX(array, row~MATCH(lookup_value1, lookup_array1, [match_type]), col~MATCH(lookup_value2, lookup_array2, [match_type]))
    - OFFSET(reference, rows, cols, [height], [width]) returns a cell/ range with specified number of rows and columns from a reference cell or range
    - CHOOSE(index_num, value1, [value2], ...) : uses index_num to return a value from the list of value arguments
- Visualization
    - Home -> Conditional Formatting -> Data Bar; Color Scales
    - Insert -> Scatter Plot: add trendline, display equation & R^2
- Scenario/Sensitivity analysis
    - Data -> What-If Analysis -> Goal Seek (set cell, to value, by changing cell)
    - Data -> Solver: set objective (max, min, value), by changing cell, add constraints (can use arrays)
    - Data -> What-If Analysis -> Data Table: select the region, link top left corner cell to the target (calculate), link row/column input to the value to change

- Home -> Analyze Data (ask AI questions)

Link: W3School - Excel Tutorial : https://www.w3schools.com/EXCEL/index.php